DOCUMENT RESUME

ED 478 989 TM 035 134

AUTHOR	Janson, David C.
TITLE	Assessment Practices of Third and Fifth Grade Science Teachers: A Comparison to the Style/Format, Process, and Content of Ohio's Proficiency Tests.
PUB DATE	2002-05-22
NOTE	123p.; Ed.D. dissertation, Ohio Wesleyan University.
PUB TYPE	Dissertations/Theses - Doctoral Dissertations (041) -- Reports - Research (143)
EDRS PRICE	EDRS Price MF01 Plus Postage. PC Not Available from EDRS
DESCRIPTORS	Educational Assessment; *Elementary School Teachers; Intermediate Grades; State Programs; *State Standards; *Student Evaluation; *Teacher Made Tests; *Test Construction; Testing Programs; Textbook Content
IDENTIFIERS	*Ohio; Ohio Fourth Grade Proficiency Test; Ohio Sixth Grade Proficiency Test

ABSTRACT

This descriptive study compared the assessment practices of 10 elementary school teachers over a 11 weeks to those of Ohio's fourth and sixth grade science Proficiency Tests. Findings show that the assessment practices of these teachers were not aligned with Ohio's proficiency test. The tests used in participants' classrooms contained a disproportionate number of items characterized as low-level in terms of their cognitive function. Classroom test items generally fell into three categories: true/false, completion, and matching. The remaining items were predominantly low-level multiple-choice items requiring simple recall of information. The teachers in this study showed heavy reliance on the packaged assessments that accompanied their adopted textbook series with little use of teacher-designed instruments. This finding differs from findings of previous researchers who reported that most teacher assessments were done with teacher-made tests. The lack of alignment between classroom tests and Ohio's Proficiency Test is a concern because previous researchers and the teachers in this study believe that aligning classroom tests with high-stakes assessment improves student performance. Other research shows teachers teach what they test suggesting that the curriculum would be better aligned with state expectations if tests were more in line with the proficiency tests. The study found that textbooks and their assessment packages are not aligned to most state standards and that teachers need help in developing better assessments. Results suggest directions school administrators might take to facilitate inservice training for current teachers and could be helpful to textbook publishers and educators on adoption committees. (Contains 5 figures and 98 references.) (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 478 989

PERMISSION TO REPRODUCE AND
DISSEMINATE THIS MATERIAL IN
MICROFICHE ONLY HAS BEEN GRANTED BY

D. Janson

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)

2B

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

ASSESSMENT PRACTICES OF THIRD AND FIFTH GRADE SCIENCE TEACHERS: A COMPARISON TO THE STYLE/FORMAT, PROCESS, AND CONTENT OF OHIO'S PROFICIENCY TESTS

A dissertation submitted to the

Division of Research and Advanced Studies
Of the University of Cincinnati

In partial fulfillment of the
Requirements for the degree of

Doctor of Education (Ed.D.)

In the Department of Curriculum and Instruction
Of the College of Education

May 22, 2002

by David C. Janson

B.A., Ohio Wesleyan University, 1973
M.Ed., Xavier University, 1982

Dr. Glenn Markle, Chairperson
Dr. Piyush Swami
Dr. Gregg Tracy
Dr. John Varis

TM 035 134

BEST COPY AVAILABLE

Abstract

This descriptive study is addressed to policy-makers, textbook publishers, teachers, principals, and curriculum directors. It compares the assessment practices of ten elementary teachers over a period of 11 weeks with Ohio's fourth and sixth grade science Proficiency Tests. Results show that the teachers' assessment practices were not aligned with Ohio's Proficiency Test. The tests used in the participants' classroom contained a disproportionate number of items characterized as low-level in terms of their cognitive function. Classroom test items generally fell into three categories - true/false, completion, and matching. The remaining items were predominantly low-level multiple-choice items requiring simple recall of information. The teachers in this study showed a heavy reliance on the packaged assessments that accompanied their adopted textbook series with little use of teacher-designed instruments. This differs from the findings of previous researchers who reported that most teacher assessments were done with teacher-made tests. The lack of alignment between classroom tests and Ohio's Proficiency Test is a concern because previous researchers and the teachers in this study believe that aligning classroom tests with high-stakes assessment improves student performance. Other research shows teachers teach what they test suggesting that the curriculum would be better aligned with State expectations if classroom tests were more in line with the proficiency tests. This study found that textbooks and their assessment packages are not aligned to most state standards and that teachers need help developing better assessments. The results of this study suggest directions school administrators might take to facilitate inservice training for current teachers and could be helpful

to textbook publishers as well as educators serving on adoption committees. Since high-stakes testing of students in the nation's public schools and school accountability seem destined to remain a part of the American educational system, educators at all levels - teachers and administrators at the local level, consultants and administrators at the state level, and policymakers at the state and national levels - may want to consider the implications of these findings.

Acknowledgements

Special thanks go to Dr. Glenn Markle. He has served in a sterling manner as chairperson of both of my committees. His guidance and wealth of experience has proven valuable to me as a student and, more importantly, a practitioner in the field. His methods as a teacher helped me to reflect on my beliefs about instruction and have changed me in ways that will benefit students during the remainder of my service as an administrator and educational leader.

I owe much gratitude to the other members of my dissertation committee who also served as members of my program of studies committee - Dr. Piyush Swami, Dr. Gregg Tracy, and Dr. John Varis. Each of these men willingly shared with me their valuable knowledge, wisdom, and unique insights into the world of education. In addition, I am very grateful for the support and counsel from Dr. Michele "Mikie" Hummel, who served on my program of studies and encouraged me throughout my experience. Her support is only exceeded by the kindness that she showed a stranger looking for the opportunity to stay connected with the "front-line troops" during his residency. For each of these individuals, the best way that I can repay them is to someday do the same for another student looking to broaden his/her knowledge and skills as an educational leader.

I am very grateful for the support of my current superintendent, Jerry Clark, and the Piqua City Board of Education who encouraged me to finish my doctoral program and facilitated its completion in a variety of ways that validates my decision to put my skills and knowledge to work for their students.

To my wife, Janet, thank you. You are a jewel. You never asked for this experience, but your love and continuous support made its completion possible.

Finally, this study is dedicated to the teachers who so willingly offered their services as participants in this study. It takes courage to participate in a descriptive study. Without the willingness of teachers to undergo such close scrutiny, it would not be possible to gather the rich kinds of data so essential to improving our noble and most important profession. Thank you.

Chapter 1 - Introduction

This introduction offers a definition of assessment, explains the purpose of the research, and offers a brief description of the Ohio Proficiency Test. It concludes with a discussion of the importance of this study.

Definition

According to McMillan (1997), classroom assessment is defined as the collection, interpretation, and uses of information to help teachers make better decisions. His conceptualization considers assessment more than testing or measurement. McMillan identifies four essential components to implementing classroom assessment -- purpose, measurement, evaluation, and use. He explains *purpose* as the reason why an assessment is performed, *measurement* as the techniques that are chosen to gather information, *evaluation* as how the results are interpreted or what criteria/performance standards are used, and *use* as how the results are used. This study proposes to focus on the measurement component of assessment.

McMillan (1997) broadens the traditional definition of the term *measurement*, a systematic process of assigning numbers to performance, "to include more qualitative ways of defining a trait, behavior, or performance, because much of what teachers do during instruction involves a less than systematic or quantitative process (p. 9)." Thus, *measurement*, as defined by McMillan,

> *"...is the process by which traits, characteristics, or behavior are differentiated. The process of differentiation can be very formal and quantitative, such as using a thermometer to measure temperature, or can consist of less formal processes, such as observation ("It's very hot today!"). Typically, measurement is used to assign numbers to describe*

attributes or characteristics of a person, object, or event, but ... qualitative information can also be used to differentiate. A test is a formalized procedure in which students respond to questions (p. 9)."

Purpose

The purpose of this research was to investigate the assessment practices especially the formal measurement practices, of third and fifth grade science teachers and compare their written tests to previous fourth and sixth grade Ohio Proficiency Tests (OPT). Specifically this research proposed to answer the following questions about teacher assessments: 1) How do the teachers' tests compare to the style and format of the questions used on the OPT? 2) How does the content assessed by the teacher test items compare to the content assessed by the OPT? 3) How do the scientific processes being tested by the teachers compare to the scientific processes assessed by the OPT? 4) Is there alignment between what the participants are doing and what and how the proficiency tests assess students?

This descriptive research study involved five third grade teachers in three different buildings and five fifth grade teachers in three different buildings. The third grade teachers were not located in the same buildings with the fifth grade teachers. This study focused on formal quantitative assessments such as teacher-designed and textbook publishers' packaged tests, usually referred to as black-line masters by teachers. Using descriptive statistics and qualitative research methods, it compares the participants' data to the fourth and sixth grade Ohio science proficiency tests. The study is limited by the size and scope of the sample because the ten participants were located in the same school district, although they were fairly evenly distributed in six different buildings across the

district. Other limitations of the study are discussed in the final chapter.

Description of the Ohio Proficiency Test

The fourth and sixth grade Ohio Proficiency Tests (OPT) were designed by the same publisher, Riverside Publishing Company. The Ohio Department of Education originally awarded the contract to develop these tests by providing Riverside and other interested bidders the same information (the *Information Guide*) used by the researcher to analyze the test items. The test assesses five different content areas - reading, writing, math, citizenship, and science.

The test developers originally constructed a bank of items from which the publisher selects different items each year. The tests are currently administered to all fourth and sixth grade students across Ohio's public schools during the month of March with exceptions only offered to special education students who are not receiving the same curriculum as other students. Students are permitted up to two and one half hours to complete each content area of the test. Prior to 2002, the five tests were administered over a period of five days with writing being administered on a Monday, reading on a Tuesday, math on a Wednesday, citizenship on a Thursday, and science on a Friday. During year 2002, Ohio's legislators mandated that the tests be administered over a two-week period with one day of no testing between each administration of the test.

The Ohio Proficiency Tests will be replaced by a series of Achievement Tests, which will be phased in over a period of several years. Current Ohio law requires that the fourth grade science be replaced by a fifth grade Achievement Test in March of 2006 and the sixth grade science test be replaced by an eighth grade Achievement Test in March of 2007.

Importance of the research

Linkage between instructional goals and assessment is generally accepted as essential to effective lesson design. For example, Tyler's (1950) linear-rational model of instructional planning suggests that teachers need to make decisions about pupil evaluation before instruction begins. Designing the assessment prior to instruction provides the goals of learning, which can then be written as behavioral objectives. Put simply, a purposeful and well-constructed (valid and reliable) assessment establishes a framework for instruction and serves as a guide for assessing the learning (Gronlund & Linn, 1990).

Burden & Byrd (1994) also link instructional goals and assessment noting that as part of instructional planning, teachers must decide whether identified instructional goals lend themselves to assessment through an affective, a cognitive, or a psychomotor domain. They suggest that although assessment of pupil knowledge pertaining to each instructional goal can be evaluated, the specific method of assessment should vary. In other words, because the goal of instruction influences the specific method of assessment that is used, different purposes will require different assessment approaches. They believe that methods selected for assessing learning should be consistent with the instructional goals. Mehrens & Lehmann (1987) share similar beliefs noting that when goals, instruction, and assessment methods are aligned, inferences about pupil knowledge are strengthened and student performance is enhanced.

Previous researchers found that teachers' assessment practices were not aligned with their instructional goals and tended to demand low levels of cognitive processing (Bol & Strage, 1996; Semb & Spencer, 1976). Therefore, it is

important to investigate the types of assessments administered in classrooms as well as the cognitive challenge associated with these types of assessments. There is also empirical evidence to suggest that the type of test items that students anticipate is related to their study strategies, which in turn affect their achievement (Crooks, 1988; Doyle, 1983; Thomas et al., 1993). Finally, many educators and policymakers believe that *what gets assessed is what gets taught* and that the format of assessment influences the format of instruction (O'Day & Smith, 1993). Given the linkage between assessment practices, instructional methodology, and student achievement, the question of what teachers' assessments emphasize and how they are designed has obvious significance with regard to increasing student performance.

In addition, large-scale assessment has been widely touted by legislators, policy makers, educators, and the general public as a lever, tool, and catalyst for school reform and improved student performance (Haertel, 1999; Kane & Khattri, 1995; Simmons & Resnick, 1993; Wolf, LeMahieu, & Eresh, 1992). New education initiatives and calls for increased accountability through the use of large-scale assessments as a necessary component of education reform have raised the demands on schools to develop more effective, integrated methods for assessing student achievement (McDonnell, 1994; Loveless, 2000; Simmons & Resnick, 1993; O'Day & Smith, 1993). Nationally, districts are being pressured to align assessment methods within their schools to state and district standards, and local curriculum and instructional practice (Cromey, 2000). In Ohio, school districts are attempting to align their written curriculum with the state's *learner outcomes* tested by Ohio's proficiency tests and the *learner competencies* of the

Ohio Graduation Test (Curriculum Alignment Project, 2000; S. Jenkins, personal interview, September 22, 2000). With the advent of high stakes testing, assessment of pupil achievement appears to be of interest to teachers and administrators anxious to align their in-house testing practices with a standards-based (i.e. high stakes test-driven) curriculum. Aligning teacher assessment with Ohio or other states' high stakes tests underscores the importance of this investigation into the assessment practices of teachers. Finally, assessment plays a critical role in the educational reform movement because tests are the yardstick used to evaluate the success of reform efforts and they directly influence classroom instruction and student learning (Bol, Nunnery, Stephenson, Mogge, 2000).

The information gained from this study is used to recommend if and what kind of specialized training is needed for preservice teachers. Additionally, it is intended that the outcomes of this study will help determine directions school administrators might take to facilitate inservice training for current teachers. Finally, information gained from this study may be helpful to textbook publishers as well as educators serving on adoption committees. Since high-stakes testing of students in the nation's public schools and school accountability seem destined to remain entrenched in the fiber of the American educational system, educators at all levels - teachers and administrators at the local level, consultants and administrators at the state level, and consultants and leaders at the national level - may want to look closely at the implications of this study.

Chapter 2 - Literature Review

This review of the literature describes how assessment is related to effective instruction and whether or not exposure to specific kinds of tests may be beneficial to students when being assessed with similar kinds of tests. The findings from previous studies that examined the assessment practices of teachers are also cited. This review concludes with a discussion of some of the methods employed by previous researchers who have studied the assessment practices of teachers.

Relationship of assessment to effective instruction and student achievement

This section discusses the importance of assessment to student achievement. It reviews some studies which suggest that using specific testing formats in the classroom can improve student success on similar formats used on mandated, high-stakes tests and concludes with a study which points to frequent assessment as a tool used by effective teachers to improve student achievement.

Research on teacher decision-making, cognitive learning, student motivation, and other topics has influenced the importance of assessment for effective teaching (McMillan, 1997). For example, two important findings are that good teachers continually assess their students relative to learning goals and adjust their instruction on the basis of this information and that assessment of students not only documents what students know and can do, assessment itself influences learning (Stenmark, 1991; Herman et al, 1992; Brophy & Alleman, 1991).

As previously mentioned, Tyler (1950) suggested that effective teachers design their assessment before planning instruction. However, prior to the nineties and the national trend of state mandated assessments, most teachers viewed assessment as something that came at the end of unit, course, or class and having little to do with instruction (Asp, 1998). With the national focus on assessment (Roeber, 1995), teachers have come to understand that assessment is both a tool for accountability and a blueprint for instruction (Linn, 1993).

Mehrens (1989) advises teachers to spend some time teaching students general test-taking skills. He notes that such skills would help students answer questions correctly if they have mastered the objectives. "Without some level of test-taking skills, even knowledgeable students could miss an item (or a set of items) because they did not understand the mechanics of taking a test (p. 3)."

There is some research suggesting support for exposing students to specific testing formats to improve their achievement on these kinds of tests. A case study of effective practices commissioned by the Ohio Department of Education and conducted by the Indiana Center for Evaluation (2001) identified those practices used by selected school districts to increase student academic achievement. Fifty Ohio school districts were selected for inclusion in the study based on their improvement in student achievement as reflected by their results on the statewide proficiency tests. Districts were asked to report on strategies that they had used to improve student achievement. District superintendents named two key staff members to identify and rank district practices in order of importance for improving student academic performance. The researchers identified test preparation as one of the prime strategies. "The teaching of test-

taking strategies, controlling testing conditions, opportunity to practice test-taking, plus seeing other techniques to provide students with familiarity with the instructions and format of the test received wide recognition.... One district used the following strategy: First thing every morning the children are given a 10-minute practice test to get them used to responding to the directions and format of the test (p 2)."

A study which compared two types of teacher-made in-class tests (multiple-choice and short answer) with a no test (control) condition to determine their relative effectiveness as aids to retention learning found that administering valid multiple-choice item and short answer item tests can improve learning weeks after the initial instruction and testing have occurred (Haynie, 1994). Similar studies found similar results (Haynie, 1990a, 1990b, 1991, in press; Nungester & Duchastel, 1982). In his study, Haynie (1994) also found evidence that multiple-choice tests promote retention learning more effectively than do short-answer tests. He...

> "...investigated the value of short-answer in-class tests as learning aids. Undergraduate students (*n*=187) in 9 technology education classes were given information booklets concerning "high-tech" materials without additional instruction. The control group was not tested initially. Students in the experimental groups were either given a multiple-choice or a short-answer in-class test when they returned the booklets. All groups were tested for delayed retention three weeks later. The delayed retention test included subtests of previously tested and new information. Both short answer and multiple-choice tests were more effective than no test in promoting delayed retention learning. No difference was found between short-answer and multiple-choice tests as learning aids on the subtest of information which had not been tested on the initial tests, however, multiple-choice tests were more effective in promotion of retention learning of the information actually contained in the immediate posttests. This study compared two types of teacher-made in-class tests (multiple-choice and short-answer) with a no test (control) condition to determine their relative effectiveness as aids to retention learning (that learning which is

> still retained weeks after the initial instruction and testing have occurred). The investigation involved instruction via self-paced texts, initial testing of learning, and delayed testing 3 weeks later. The delayed tests, which included both previously tested information and novel information that had not been previously tested, provided the experimental data for the study (p 1)."

Haynie surmised that moving information from short term to long term memory is aided by rehearsal. He concluded that multiple-choice test items are a more effective form of rehearsal than short answer test items. These studies' findings are significant because they differ from the previously held belief that students study more efficiently for essay-type examinations than for multiple-choice tests (McKeachie, 1986). McKeachie reasoned that students preparing for essay tests focus on broad issues, general concepts, and interrelationships rather than on specific details, and that studying for essay tests results in somewhat better student performance regardless of the type of exam they are given.

Smith (2000) conducted a qualitative school-wide action study in which systematic strategies were used to substantially improve student scores on mandated testing. The elementary staff instituted changes that included regularly exposing students to the test format that they would see on the standardized tests. The staff considered the teaching and practice of test-taking skills to be a basic element of their school wide instructional program to raise standardized test scores. They provided time for students to learn and practice test-taking skills in the context of teaching the school curriculum in ways that improved their academic achievement. Part of the staff's commitment included becoming familiar with the basic design, structure, and content of the standardized test that their students were required to take. They also adopted the belief that instruction

and student mastery of subject content and skills to be examined on a test are fundamental to students scoring well on any test. It became a fundamental tenet of their system of accountability that curriculum, instruction, and assessment should be aligned.

During the study, Smith (2000) noted that some teachers were initially concerned that teaching students test-taking skills and motivating them to try hard on standardized tests might be considered *teaching to the test* and that efforts to help students master important content-area skills would suffer. However, Smith found that as test scores increased, teachers saw that student learning of curricular content and skills increased greatly. Smith concluded that the strategy used by the teachers equipped students with skills that allowed them to accurately demonstrate what they had learned.

Black and William (1998) conducted a survey of the research literature seeking evidence that improving *formative assessment* can improve standards for learning. The authors defined this form of assessment as all of the activities undertaken by teachers to be used as feedback to modify teaching and learning activities. Their review of the literature found evidence that teacher's everyday assessment practices are beset with problems and shortcomings. Generally, the studies reviewed found that most tests used by teachers encourage rote and superficial learning even when teachers say they want to develop understanding. Some of the studies noted that teachers seem unaware of the inconsistency between what they state about the purpose of their assessments and what they practice. The questions and methods teachers used are not critically reviewed in relation to what they actually assess. For primary teachers, they found evidence

in much of the research that there is a tendency to emphasize quantity and presentation of work and to neglect its quality and intent. Based on their review of the research literature, they concluded that improvements in formative assessment can result in improvements in achievement.

There is also considerable evidence from a variety of sources suggesting that practice with a particular format or style of testing can be beneficial to performance on similar kinds of tests. In their review of the research for Chicago Public Schools, Borger et al (1996) found that students who are test-wise are more successful in testing situations because they are familiar with item formats and can apply test-taking strategies to solve different kinds of problems. In their interview and focus group sessions, students told the interviewers for Chicago Public Schools that they do not feel prepared to take tests. Their research led them to the following conclusions.

1. Test preparation needs to be an ongoing activity, not something that is addressed right before the spring testing period.
2. Test preparation should be part of regular classroom instruction and integrated into all subject areas.
3. Some type of assessment preparation should be utilized in every classroom on a daily basis and should be part of instruction that addresses the skills being measured. For example, teacher-made materials, worksheets, tests and oral questions can be adapted to address important test-taking skills as well as the instructional standards and critical thinking skills. Likewise, instructions to accompany materials can be written to mirror more closely the testing situations the students will encounter.

Borger et al advise teachers that because "most standardized test items require students to use critical thinking skills, it stands to reason that the more practice students get in applying their thinking skills, the better prepared they will be to take standardized tests (p 21)."

In his analysis of Virginia's Standards of Learning (SOL), Hanny (2001) advises teachers to align their curriculum, teaching, and assessment to the SOL and explains how Virginia's system of tests assesses the SOL. He notes that, "State standardized test items will be multiple choice test items. This means that if you have been using other types of items, such as matching, short answer (completion), fill in the blank, you need to think about the fact that multiple-choice items dominate these tests. Perhaps the best advice is to try and use multiple choice when measuring knowledge, comprehension, and to a degree, application, in your classroom when assessing SOL-based material." Hanny's suggestion has similar implications for Ohio's teachers. Because the fourth and sixth grade Ohio Proficiency Test for science is also predominantly multiple-choice items, Ohio's teachers should consider replacing their true/false, matching, and completion items with appropriate multiple-choice items. Indeed textbook publishers who often want to sell their product on the basis that their products are aligned to state standards and tests would be wise to take notice of this implication. Most publishers provide a set of black-line masters (a booklet of prepackaged test questions and/or tests that teachers are permitted to copy) for teachers to assess their students. If these assessment packages are not aligned to the state's system of testing, adoption of the materials becomes a catalyst for alignment issues.

Finally, a study conducted by the Educational Testing Service that linked "effective teaching methods" to students' results on mathematics and science subtests from the 1996 National Assessment of Educational Progress (NAEP) showed that students who frequently take tests outperform those frequently using on-going forms of assessment such as portfolios (Wenglinsky, 2002). The researchers looked at 15,000 NAEP scores (7,146 eighth-grade students who took the mathematics assessment and 7,776 eighth-grade students who took the science assessment). They related these scores to questionnaires that were completed by student test-takers and their teachers and principals. The researchers wanted to investigate if what teachers did in their classrooms had an effect on student performance. They collected data on classroom practices including assessment practices. For both math and science, the researchers found that students who took point-in-time tests performed higher than those whose teachers used portfolios and projects as assessment (i.e., 92% of a grade level in science).

In summary, there is some evidence that assessment practices positively affect student learning. For example, effective teachers use frequent assessment. Several studies suggest that the methods teachers use to assess student progress are linked to improvements in student achievement. Also, tests given at a particular point in time for math and science instruction are associated with higher student performance than on-going techniques, such as portfolio or project-based assessment. Other studies provide evidence that students who are test-wise are more successful in testing situations because they are familiar with item formats and can apply test-taking strategies to solve different kinds of

problems. For primary teachers, researchers found evidence that there is a tendency to emphasize quantity and presentation of work and to neglect its quality and intent. Researchers conclude that improvements in assessment methods can result in improvements in achievement.

Assessment practices of teachers

This section reviews the assessment practices exposed by various studies. It cites previous studies which indicate that teachers do not follow many of the practices recommended by measurement experts and concludes with citation of earlier studies which show most of the evaluation done in schools is done with teacher-made tests.

Prior studies show a wide variation in teachers' use of recommended measurement practices when assessing pupil achievement (Brookhart, 1994; Frary et al., 1993; Gullickson & Ellwein, 1985; Stiggins et al., 1989). Studies show that teachers and administrators usually do not receive training on how to assess students or how to apply assessment information to instruction (Cizek, 2000; Scharfer & Lissitz, 1987; Wise, Lukin, & Roos, 1991). Other studies show that most teacher tests rely too heavily on students' recall of information and generally do not follow measurement practices recommended by experts (Milton, Pollio and Eison, 1986; Applebee, 1981, 1993; Bloom, 1984; Goodlad, 1984; Marso & Pigge, 1991; Squire & Applebee, 1968).

A study using document analysis and ethnographic research found a significant proportion of the assessment materials in a Grade 10 English Course did not involve written response. While Kahn (2000) found a greater use of extended written responses than reported in previous studies (Bloom, 1984;

Goodlad, 1984; Marso & Pigge, 1991; McNeil, 1986; Squire & Applebee, 1968), most of the students' summative assessments were based on lower level questioning strategies. In the 950 materials analyzed by Kahn, 65% of the total number of points for the semester involved questions that were multiple-choice, matching, and true/false while thirty-five percent of the total number of points for the semester involved written and oral responses. Kahn found no short-answer or fill-in-the-blank items. The final examination was multiple-choice and matching without any written responses. During informal conversations, the teachers in Kahn's study...

> "*...indicated that they used a certain number of multiple-choice items because of the number of students they had in their classes. Participants felt that number of students that they taught each day (four to five classes each semester with a total of 120 to 135 students) made it impossible to give essay tests instead of multiple-choice tests on a regular basis. Also, because grades had to be submitted on the last day of final examinations, they said they reported using a multiple-choice format on the examination to complete the grading in a timely manner* (p. 7)."

Researchers who have examined tests and other materials developed by teachers to assess student learning in their classrooms reported a dominance of short answers and recall of information with little emphasis on writing or encouraging students to develop and defend their own interpretations. The tests and assignments examined in these studies tended to reflect a traditional model of teaching as transmission of knowledge (Applebee, 1981, 1993; Bloom, 1984; Goodlad, 1984; Marso & Pigge, 1991; Squire & Applebee, 1968). Most of the studies reported on sample tests collected from teachers representing a range of schools, grade levels, and courses. These studies, however, do not reveal much about the kinds of assessment that teachers use over a period of time such as a

quarter or semester.

Finally, with regard to the use of teacher-made tests versus publisher's tests, earlier studies show most of the evaluation done in schools is done with teacher-made tests (Haynie, 1983, 1991, 1992; Herman & Dorr-Bremme, 1982; Mehrens, 1987; Mehrens & Lehmann, 1987; Newman & Stallings, 1982). The previous finding by these researchers is interesting because it differs from the finding of this study of ten elementary teachers' assessment practices in science.

In summary, previous studies indicate that teachers do not follow many of the practices recommended by measurement experts such as McMillan (1997) and Wilson (2002). Teachers and administrators typically do not receive training on how to assess students or how to apply assessment information to instruction. There is also evidence that most teacher tests rely too heavily on students' recall of information with short answers and recall of information dominating the frequency of items. Evidence suggests that there is generally little emphasis on writing a paragraph or more on encouraging students to develop and defend their own interpretations. Finally, of special interest to the researcher, previous studies show teachers prefer to design their assessments rather than rely on publisher's tests.

Methods employed by previous researchers

This section describes the methods used by previous researchers looking into the assessment practices of teachers. Most recent studies conducted on teacher assessment practices have been descriptive case studies (Cromey, 2000; Campbell & Evans, 2000; Kahn, 2000). These studies focused on exposing the practices of teachers or their ability to interpret and use assessment data.

Cromey (2000) gleaned information from the case studies of nine Michigan public schools (five elementary, three middle, and one high school) which were conducted by the North Central Regional Educational Laboratory (NCREL) in 1999. Schools were selected for the study based on nominations made by Michigan State Board of Education staff. Selected schools were located in different regions of the state and were diverse in terms of student enrollment figures and key student characteristics (e.g. socio-economic status, ethnicity). NCREL researchers performed on-site, semi-structured interviews related to school-based assessment practices with teachers, principals, and district-level staff. A total of 46 interviews were conducted in nine sites. Teachers were selected for participation by their curriculum director or principal based on their prior experience with using student assessments and data. Participants were also asked to provide documentation that could help researchers better understand the student assessments being developed and pursued locally, and information about local efforts to set standards and align curriculum. The researcher found that the educators in this study had little formal training in student assessment and the use of related data to improve teaching and learning and the educators lacked a process for using data that supports ongoing, continuous improvement in their schools. Also, noting that some of the educators seemed to struggle with increasing levels of assessment-linked accountability and the community attention to assessment data, the researcher found that they viewed the results from the mandated assessments as punitive rather than vehicles for school improvement, which, in turn, lead to a perceived lack of face validity with the large-scale assessments administered to their students. She

concluded that for educators, having a foundation in data analysis and assessments is necessary, but not sufficient for them to effectively synthesize assessment data at the school level. She also recommended that all state-mandated assessments be aligned to clearly defined learning standards.

Campbell and Evans (2000) conducted a case study of 65 student teachers' classroom assessments. The researchers created rubrics to ensure that their critique of preservice teachers' assessment methods were valid and reliable. They structured the rubric as a modified checklist to identify (a) plan for assessment; (b) method of assessment (e.g., paper and pencil, alternative); (c) learning goals/instructional objectives; (d) degree of match (i.e., curriculum alignment) between learning goals/objectives and selected assessment(s); and (e) inclusion of a scoring rubric to ensure consistent scoring of constructed performance behavior. The extent to which the above criteria were found provided evidence to support the reliability, validity, and appropriateness of the preservice teachers' assessment methods. The authors found that, during student teaching, preservice teachers do not follow many of the assessment practices recommended in their coursework.

Another study combined mostly document analysis with some interview and ethnographic data. Kahn (2000) examined materials designed by a group of teachers to evaluate student learning and instructional effectiveness during the first semester of a Grade 10 English course at a large suburban high school in the Midwest. The participants provided copies of all the major tests and other materials, as well as copies of the final examination that was used during the semester for assessment purposes. The researcher also collected other

materials used in the unit - copies of handouts for students (e.g., study guides, writing assignments), quizzes, and tests. The materials that the researcher analyzed included all the tests on the six units used by the teachers during the semester, as well as a final examination. She also collected one writing assignment and one speech assignment.

Kahn used a point system to analyze the materials. Multiple-choice tests, essay questions, written compositions, and speeches were assigned point values to reflect their relative weight in assigning grades to students. The point values were considered when the materials were analyzed. Data were prepared for analysis by coding the test items according to the type of response that students were asked to provide -- multiple-choice, matching, true/false; short answer (a word or brief phrase); or open-ended written or oral response (requiring at least 1-2 sentences). In addition, the researcher coded materials to indicate whether they involved direct or indirect means of assessment. Kahn analyzed the data by looking for patterns and findings that seemed important.

The researcher also talked informally with the participants about their instruction, tests, and perceptions of student learning in their classrooms. She attended meetings at which the teachers were present and during which they discussed their views about teaching and assessing learning. Relevant comments from the conversations and discussions were recorded as notes by the researcher. In the process of analyzing the data, she noted questions that arose about the materials and how they were used, which she later asked the teachers to address in informal conversations. Kahn found that, in the set of materials analyzed, 65% of the total number of the students' points for the

semester involved questions that were multiple-choice, matching, and true/false. She also found no short-answer (requiring only a word or brief phrase) or fill-in-the-blank items. Thirty-five percent of the total number of points for the semester involved written and oral responses--usually multi-paragraph compositions.

Other studies used questionnaires and teacher focus group interviews. One such study investigated changes in teachers' assessment practices (Bol, Nunnery, Stephenson, Mogge, 2000).

Finally, Frary, Cross, and Weber (1993) administered a 44 item questionnaire to 536 randomly chosen secondary teachers from Virginia's public schools to determine their beliefs and practices concerning testing and grading. A factor analysis of the opinion responses identified seven factors, which were interpreted as representing the extent to which the responders agreed. Two of the factors identified in the study are of particular interest to the researcher - 1) multiple-choice tests are generally undesirable and 2) difficult tests are pedagogically unsound. Both factors were contrary to what Frary, Cross, and Weber recommended.

The studies reviewed in this section generally used some combination of methods including questionnaires, document analysis, and interviews to expose the assessment practices of teachers on a variety of levels in a variety of content areas. Most of the studies were descriptive or ethnographic. Findings from these prior studies relevant to this investigation of teachers' science assessment practices include: 1) educators receive insufficient training in student assessment and the use of related data to improve teaching and learning; 2) large-scale assessments may lack credibility with some educators because they don't

consider them valid measurements of what actually occurs in the classroom; 3) some educators see the attention to large-scale assessment results as punitive; 4) preservice teachers do not follow the assessment practices recommended in their coursework; 5) there is some evidence that teachers perceive construction of multiple-choice items to be difficult and believe that multiple-choice items are pedagogically weak.

Chapter 3 - Methods

This study investigates the assessment practices of third and fifth grade science teachers and, specifically, examines their construction and use of written tests and how they compare to the format and style of the questions on the fourth and sixth grade Ohio Proficiency Tests. In addition, the participants' tests are compared to what the Ohio Proficiency Tests are assessing in terms of content and level of cognitive challenge in the form of scientific processes. This exploratory and descriptive study combines interview research (Patton, 1990), document analysis (Schmidt & McKnight, 1995; Porter, 1995; Patton, 1990), and survey research (Cozby, 1997; Patton, 1990).

Participants

The ten participants for this study were selected from the third and fifth grades of a small city school district in Ohio of approximately 4000 students. Five teachers were selected from each of the two levels. Approximately 35% of the district's students receive free and reduced lunch subsidies. Analysis of past test scores indicated the fourth and sixth grade students scored below the state average and well below Ohio's established benchmark for effective schools. The participants' range of teaching experience included seven participants with twenty-five or more years, one participant with twenty-two years, and two participants with less than seven years. The two teachers in this latter group were in their first year of regular education. Their previous experience was in the area of special education. All of the participants were female and certified to teach elementary students.

Data Sources

Data were collected from three separate sources - documents, interviews, and questionnaires. First, teachers were asked to submit blank copies (without student names or responses) of all written tests, quizzes, assignments, and projects that were used to formally assess student mastery. For the purpose of this study, formal assessment was defined to the participants as a paper and pencil test/assignment or a project-based assessment using a rubric similar to the rubric shown in Appendix P and recorded as a grade. This rubric was taken from the *Information Guide for Science* from the Ohio Department of Education. Next, the ten participants were interviewed for approximately one hour. These interviews were audio-recorded and later transcribed for analysis. Following her interview, each teacher was given a questionnaire to complete later.

Document sample

The ten teachers in this study all submitted hard copies of the assessments that they used during the first quarter of the 2001-2002 school year. Assessments included any test/assignment or project that was graded. The teachers were also asked to submit any rubrics that were used to assess short or extended response items and projects. Each item of the documents was analyzed and sorted according to style/format, scientific process, and content.

Descriptors for style and format fell into three main categories: multiple choice, short answer, and extended answer. The multiple-choice items were further arranged according to three subsets: a simple question that involved one sentence or could be answered with only one sentence of information, a passage of text involving multiple sentences that were necessary for the responder to read

prior to identifying the correct response, and figures or tables that required analysis or interpretation to identify the correct response.

The short and extended answer items were arranged according to descriptions taken from the Ohio Department of Education's Information Guide for the fourth and sixth grade proficiency tests.

To analyze the short answer items of the third grade teachers, the fourth grade descriptors from the short answer items of the Information Guide were used as follows.

1 *Provide an explanation.*
2 *Interpret a figure, table, or text.*
3 *Explain a cause and effect; provide examples.*
4 *Explain a conclusion.*

To analyze the short answer items of the fifth grade teachers, the sixth grade descriptors from the short answer items of the Information Guide were used. The categories were as follows.

1 *Provide an explanation for an answer.*
2 *Make & justify predictions.*
3 *Propose a procedure to resolve an investigation.*
4 *Interpret a figure, table, text or graph.*
5 *Explain a cause and effect; provide examples.*
6 *Explain cause of certain natural phenomena.*

Descriptors for extended answers were taken from the Information Guide for the fourth grade test and used to analyze the extended responses of the third grade teachers. The descriptors were as follows.

1 Evaluate *the effectiveness of a procedure.*
2 *Make predictions and explain reasons.*
3 *Make and justify conclusions.*
4 *Describe natural processes or interactions among the components of a biological or physical system.*
5 *Create and use classification systems.*

Descriptors for extended answers were taken from Information Guide for the sixth grade test and classified into one of the following categories for the fifth grade.

1 *Evaluate a procedure and point out its flaws or suggest improvements.*
2 *Make and justify predictions or conclusions.*
3 *Describe natural processes or interactions among the components of a biological or physical system.*
4 *Make or use inferences in resolving an investigation.*
5 *Interpret information from figure, graph, paragraph, or drawing.*

The Information Guide provided by ODE also describes three scientific processes: *acquiring scientific knowledge*, *processing scientific knowledge*, and *extending scientific knowledge*. According to the Information Guide, each of these categories can be arranged into smaller categories. Consultants from ODE who were contacted by the researcher refer to these smaller categories as "*sets of grains*" (colloquial terminology used by ODE consultants to categorize the processes being assessed). The researcher used the *grains* as descriptors to sort each of the items from the OPT and the participants documents.

The *first set of grains* for the category of acquiring scientific knowledge consisted of the same processes for both the fourth and sixth grade tests:

1 *recalling,*
2 *observing,*
3 *collecting, and*
4 *recording data and information from a variety of sources and representations and performing operations and making measurements to obtain data.*

The *second set of grains* for the category of acquiring scientific knowledge consisted of these processes for the third grade:

1 *take measurements from instruments,*
2 *read graphs and tables,* and
3 *recall fundamental facts, concepts or relationships.*

The *second set of grains* for the category of acquiring scientific knowledge consisted of these processes for the fifth grade:

1 *take measurements and/or make observations or identifications from graphs, keys, tables, or drawings;*
2 *read graphs and tables;*
3 *recall fundamental facts, concepts or relationships.*

The *first set of grains* for the category of processing scientific knowledge consisted of the same processes for both grades:

1 organizing,
2 interpreting,
3 manipulating,
4 verifying,
5 summarizing, and
6 *reformulating observations and data.*

The *second set of grains* for the category of processing scientific knowledge consisted of these processes for the third grade:

1 interpret data or information form text, graphs, tables or drawings;
2 recognize relationships, structure, and function among objects or organisms;
3 use simple identification keys or classification systems;
4 *recognize procedures appropriate to given investigation.*

The *second set of grains* for the category of processing scientific knowledge consisted of these processes for the fifth grade:

1 interpret data or information form text, graphs, tables or drawings;
2 recognize or infer relationships, structure, and function among objects or organisms;
3 *recognize procedures appropriate to a given investigation.*

The *first set of grains* for the category of extending scientific knowledge consisted of the same processes for both grades:

1 evaluating,
2 applying,
3 formulating,
4 transforming, and

5 *communicating ideas and hypotheses in various contexts.*

The *second set of grains* for the category of extending scientific knowledge consisted of the same processes for both grades:

1 *propose solutions or conclusions based on data from graphs, tables, text, or drawings;*
2 *determine and/or explain the effectiveness of a procedure;*
3 *use spatial, inductive, or deductive reasoning in problem solving.*

Interview

The audiotapes from all ten of the teacher interviews were analyzed for this study. The interviews were conducted individually following the document phase of this study and were completed during December of 2001. Inductive analysis of the transcribed interviews was conducted using Folio Views (version 5.0). Units of analysis were based on the interview guide (described in the next paragraph) and fell into two main categories-beliefs and stated practices. Each of these categories was analyzed for emergent themes or patterns.

Interviews were limited to one hour. Generally, these questions were used during the interview as a guide, but the researcher abandoned them when necessary to pursue important revelations by probing specific responses to the initial questions: 1) *How do your paper and pencil test items compare to those of the OPT?* 2) *How well aligned are the publisher's tests to the style and format of the OPT?* 3) *How well is your assessment of student learning linked to your instructional objectives?* 4) *How well are your district's learner objectives aligned to the learning outcomes assessed by the OPT?* 5) *To what degree do you pattern your written tests after the OPT in terms of style/format?* 6) *How do you feel about the OPT? Is it testing the right things?* 7) *Can student performance on*

the OPT be enhanced by using the same format with your classroom assessments? 8) *Is there anything else you would like to share with me about your assessment practices or the OPT?*

Teacher questionnaire

A two-part questionnaire (see appendices) was developed to measure teachers' self-perceived assessment practices. In Part I of the questionnaire, teachers were asked to describe their level of instruction and assessment by choosing which one of six rubric categories best applied to each of several *learning outcomes*. The outcomes were taken from the Oho Department of Education's *Guidelines for the fourth and sixth grade proficiency tests*. The questionnaire for the third grade teachers used the 19 learning outcomes assessed by the fourth grade test and the questionnaire for the fifth grade teachers used the 17 learning outcomes assessed by the sixth grade test. The rubric, shown on the questionnaires (Appendix M), was structured as a modified checklist with the following elements: (a) not relevant to this grade level or not addressed by the schools' printed or taught curriculum, (b) minimal instruction with no formal assessment, (c) instruction with informal assessment using checklists or teacher observation, (d) instruction with formal assessment using publisher's black-line masters or tests, (e) instruction with formal assessment using teacher-designed tests with T-F, matching, simple multiple choice, and (f) instruction with formal assessment using teacher-designed tests/assignments or projects aligned to the style/format of the OPT. For the purpose of this study, formal assessment was defined to the participants as a *paper and pencil test* or a *project with a rubric* similar to the rubrics used to assess extended and short

responses on the OPT. A copy of the rubric used by ODE to assess extended responses on the OPT was given to each participant with their questionnaire.

Part II of the questionnaire consisted of three questions. Teachers were first asked to estimate what percentage of their first quarter evaluations (those assessments which were entered in the teacher's grade book) were used to base their evaluation of student progress for each of the following categories: *teacher-designed tests and quizzes*, *publisher-designed tests and quizzes*, *student participation in classroom activities* (discussions, cooperative learning, projects, independent practice), *oral exams*, *group projects*, *individual projects*, *portfolios*, *homework assignments*, or *other*. If they selected *other*, the participants were asked to provide an explanation. Participants were instructed to place a "0" next to any item that was not used to formally evaluate students. The second question asked teachers to consider the same categories from the previous question for all four quarters and indicate which items would differ and how they would differ. The third question asked the teachers to indicate on a forced answer format how frequently they attempt to mimic the style of the Ohio Proficiency tests when they write their own tests and quizzes. Choices for the third question were: less than 10 % of the time, 10 - 19% of the time, 20 - 29% of the time, 30 - 39% of the time, 40 - 49% of the time, 50 - 59% of the time, 60 - 69% of the time, 70 - 79% of the time, 80 - 89% of the time, and more than 90 % of the time.

The questionnaires were provided immediately following each teacher's interview. Each teacher was given an overview of the questions and the researcher's' definitions of such terminology as *formal assessment*, *project with rubric aligned to the OPT rubrics,* and *alignment with the style and format of the*

OPT. For this study, *formal assessment* was defined as a paper and pencil test/assignment or a project with assessment using a rubric similar to the rubrics found in the OPT Information Guide for science (Ohio Department of Education). *Alignment with the style and format of the OPT* was defined as *using multiple-choice items that are usually preceded by text, graphs, drawings, or figures requiring the student to do one of the following:*

- *recall information,*
- *observe phenomenon,*
- *collect information and record data and information from a variety of sources and representations and perform operations and make measurements to obtain data,*
- *interpret data or information from text, graphs, tables or drawings,*
- *recognize relationships, structure, and function among objects or organisms,*
- *use simple identification keys or classification systems,* or
- *recognize procedures appropriate to a given investigation,*

and/or *using short answer and extended answer items that are always accompanied or preceded by text, graph, figures, or drawings that ask the student to do one of the following:*

- *interpret data or information from text, graphs, tables or drawings,*
- *recognize relationships, structure, and function among objects or organisms,*
- *use simple identification keys or classification systems,*
- *recognize procedures appropriate to a given investigation,*
- *provide explanation,*
- *explain cause-effect,*
- *interpret figure or provide examples or explain conclusion,*
- *evaluate effectiveness of procedure,*
- *make predictions & explain reasons,*
- *make & justify conclusions,*
- *describe natural processes,*
- *interactions among the components of a biological or physical system,*
- *create and use classification systems.*

In addition, the teachers were given an opportunity to seek clarification regarding any of the questionnaire items and related terminology before leaving

with the questionnaire. Teachers were instructed to return the questionnaires to the researcher in a sealed envelope marked as confidential within a week and not to spend more than 30 minutes completing it.

Reliability of the questionnaire was not statistically verified, but the questionnaire was divided into multiple sections to gather the same kinds of data through different types of questions. Face validity was established by overlaying the research questions with the questionnaire.

Data Analysis

Triangulation (Patton, 1990) was conducted by comparing data from the document analysis and the survey data with analysis of the interview data. Analysis of the transcribed interviews was conducted using Folio Views (version 5.0). Units of analysis were determined by emergent themes and categorized according to commonality.

Each item from the printed tests, quizzes and other assessment documents collected by the researcher was coded for format according to its type of assessment. The type of assessment was based on the Ohio Department of Education's description of recommended assessments for measuring student competency (1995). This document describes *Type I-A* as items that are multiple-choice and *Type I-B* as multiple-choice items that follow a reading passage or reference a figure or graph that must first be interpreted. *Type II* includes short answer items and *Type III* includes extended answer items. Items will also be coded for *Traditional* (true-false, matching, fill-in, completion) or *Alternative* (portfolio, thematic-based project, etc). The percentage of each of these types was compared to an analysis of past fourth and sixth grade Ohio

Proficiency Tests (years 2000 and 2001). Teacher test items were coded for content and the type of scientific process that was assessed using the descriptors provided by the Information Guide for Proficiency Tests from the Ohio Department of Education. All of the document and questionnaire data were recorded onto spreadsheets using Microsoft Excel from the Office 2000 package and displayed as simple descriptive statistics for analysis. The same method of analysis was used with the data from the Ohio Proficiency Tests.

To determine the category of scientific process for the OPT items, the researcher consulted the *Answer Key* (2000, 2001) for each of the appropriate tests provided on-line by the Ohio Department of Education. These documents provide a matrix of identifiers for each item of the OPT. Included in the matrix are the *item number*, the content area subscale or *strand*, the level of understanding or *scientific process*, the content *learning outcome number*, a special code known only to ODE, and the correct answer.

In this investigation, a considerable percentage of the participants' items were found to be in the category of *acquiring scientific knowledge*. Therefore, only the category of acquiring scientific knowledge was analyzed according to the *grains* for the OPT and the teachers' items. The limitation of the process used by the researcher to classify the items according to the *grains* of scientific process is discussed in the final chapter.

Finally, to determine which learning outcome the participants' items were assessing, the researcher used the descriptions of the learning outcomes provided by ODE. While this information was available in the Answer Key for the OPT items, researcher's judgment was used to determine which learning

outcome, if any, the participants' items were assessing. By conducting the analysis of the OPT first, the researcher was able to see precise models for each of the *learning outcomes*. The limitation of this process is discussed in the final chapter.

Chapter 4 - Results

The first half of this chapter focuses on the fourth and sixth grade science OPT. Before coding and analyzing the items from the participants' documents, the investigator analyzed the OPT test items. To improve reliability of the results, the items from two different years of the OPT (years 2000 and 2001) were studied and categorized. The items from the test were analyzed according to their style and format, their scientific process, and their content. The results were recorded onto spreadsheets for descriptive analyses. This work enabled the investigator to compare the data from the participants to the OPT. The results of the analyses of the OPT items are shown in the expanded tables of appendices A - F. The figures shown in the next section of this chapter graphically illustrate key findings from the analyses of these data.

The second half of this chapter is devoted to describing results from the teacher data. It consists of four sections - document results for content; document results for scientific process; document results for use of black-line masters, teacher-designed tests, and rubrics; document results for format and

style; teacher questionnaire results; teacher interview results. It concludes with a summary of the primary and secondary findings from analyses of the OPT and participants' data. Ancillary findings are also discussed.

Figure 1 shows that the majority of the sixth grade items from a composite of years 2000 and 2001 of the science OPT were multiple-choice items. Of that category, those multiple-choice items requiring students to *process information from a figure, graph, or table* prior to choosing the correct response ranked highest. However, because the Ohio Department of Education assigns different values to the different kinds of items, an analysis of these values was also conducted to determine if the pattern would differ significantly.

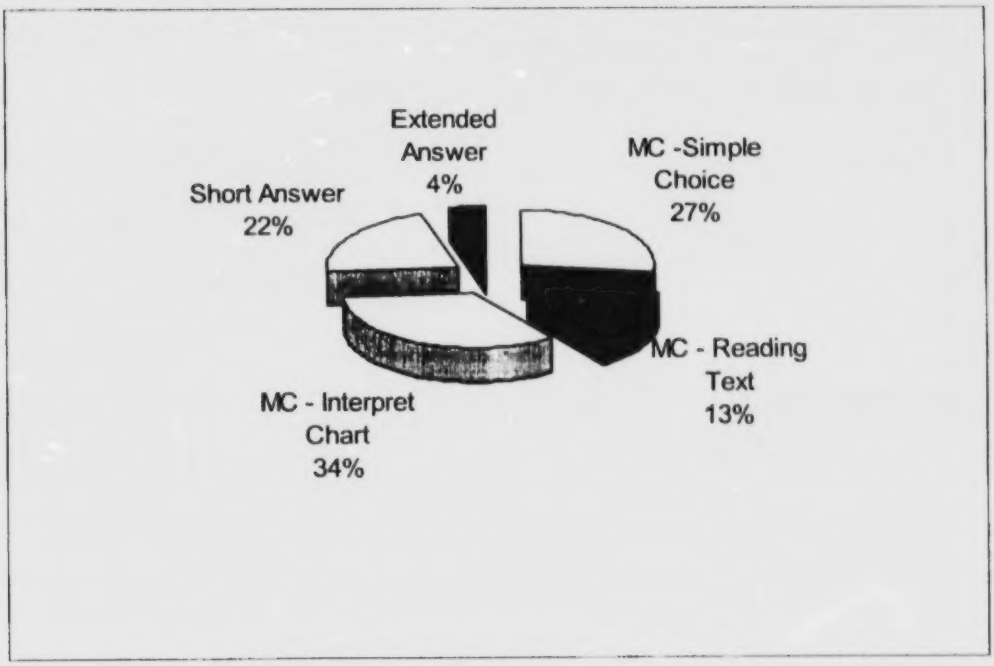


Figure 1 Percentage of Various Types of Questions, 6th Grade OPT

Figure 2 shows the value of each of these same items as a percentage of

the total raw score. The raw score is determined using a system of weights assigned by the ODE. All multiple-choice items are valued at one point each, all short answer items are valued at two points each, and all extended response items are valued at four points each. In this analysis, the *short answer items* ranked highest with *multiple-choice items requiring students to process information* ranking second. However, combining all three types of multiple-choice items placed it as the highest-ranking category in both analyses. All of the multiple-choice items made up 74% of the total items. Figure 2 shows that all multiple-choice items were worth 55% of the possible total points. The relative importance of the various kinds of items differs somewhat, but the pattern found in the first figure relative to the overall importance of the multiple-choice items remains the same in this second figure. Multiple-choice items are clearly an important part of the sixth grade science OPT.

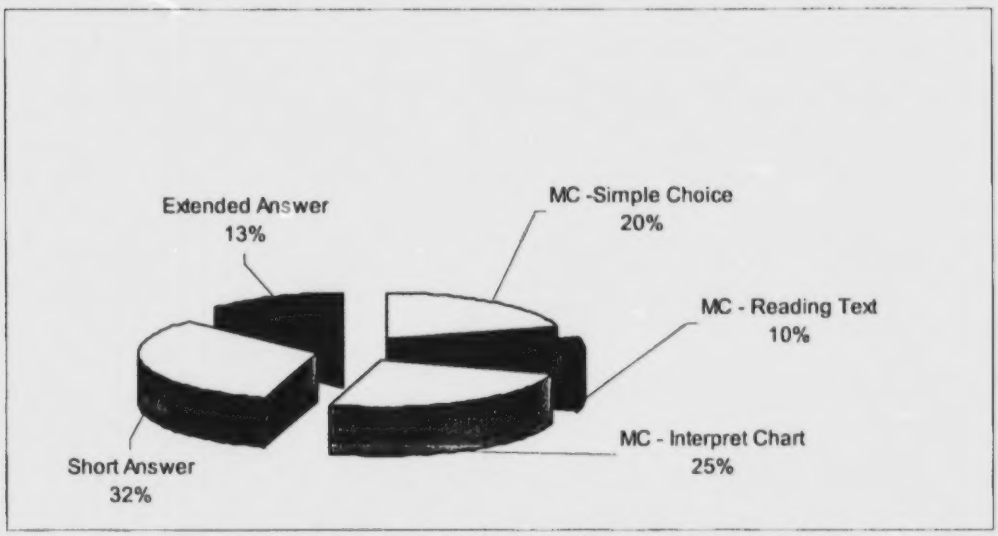


Figure 2 Percentage of Types of Questions Based on Weighted Scores, 6th OPT

Similar analyses of the fourth grade OPT were conducted and are displayed in Figures 3 and 4. With regard to the format and style of the items, a pattern similar to the sixth grade test emerged with 34% of the multiple-choice items requiring the *processing of information from charts, graphs, or tables*. This kind of multiple-choice item clearly dominates all other types of items.

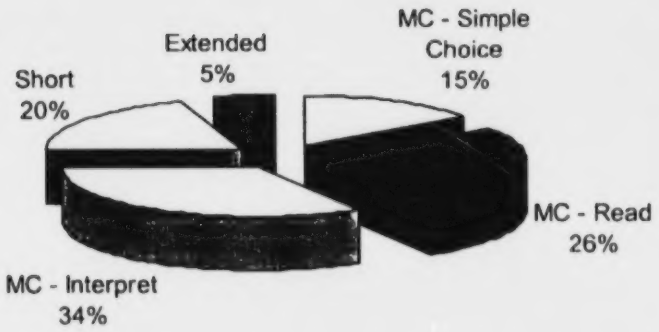


Figure 3 Percentage of Various Types of Questions, 4th Grade OPT

Figure 4 shows the value of each of above items as a percentage of the

total raw score. The raw score is determined using a system of weights assigned by the ODE. All multiple-choice items are valued at one point each, all short answer items are valued at two points each, and all extended response items are valued at four points each. In this analysis, the *short answer items* ranked highest with *multiple-choice items requiring students to process information* ranking second. However, combining all three types of multiple-choice items placed it as the highest-ranking category in both analyses. Figure 3 shows that all multiple-choice items made up 75% of the total items. Figure 4 shows that all multiple-choice items were worth 55% of the possible total points. Multiple-choice items are also clearly an important part of the fourth grade science OPT.

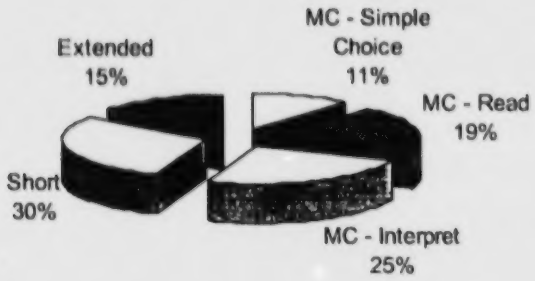


Figure 4 Percentage of Types of Questions based on Weighted Scores, 4th OPT

The sixth grade OPT assesses 17 different learning outcomes that are grouped into four different strands: *nature of science*, *physical science*, *earth and space science*, and *life science*. Overall, the multiple-choice items dominated the total number of items across 15 of the 17 learning outcomes. Learning Outcomes #1, 3, 4, 5, 8, 10, 12, and 17 favored the multiple-choice items over the short and extended answer items at better than a 3:1 ratio. Of the 92 total items for years 2000 and 2001, 34% of them were multiple-choice items that

required students to interpret a figure, table, or graph prior to responding. Of the 68 multiple-choice items, almost half (46%) required students to work with a figure, table, or graph. Clearly, students need to be competent with the use of such tools to be successful with the OPT science.

Because the participants in this study were teaching primarily in the *life science* strand during the period of time in which their assessment documents were collected by the researcher, the analysis of this strand is depicted in Figure 5. The researcher found that the *life science* strand does not follow the general overall pattern of distribution with regard to the format and style of the test items. In this strand, the dominant type of item is clearly *simple multiple-choice* followed by *short response*.

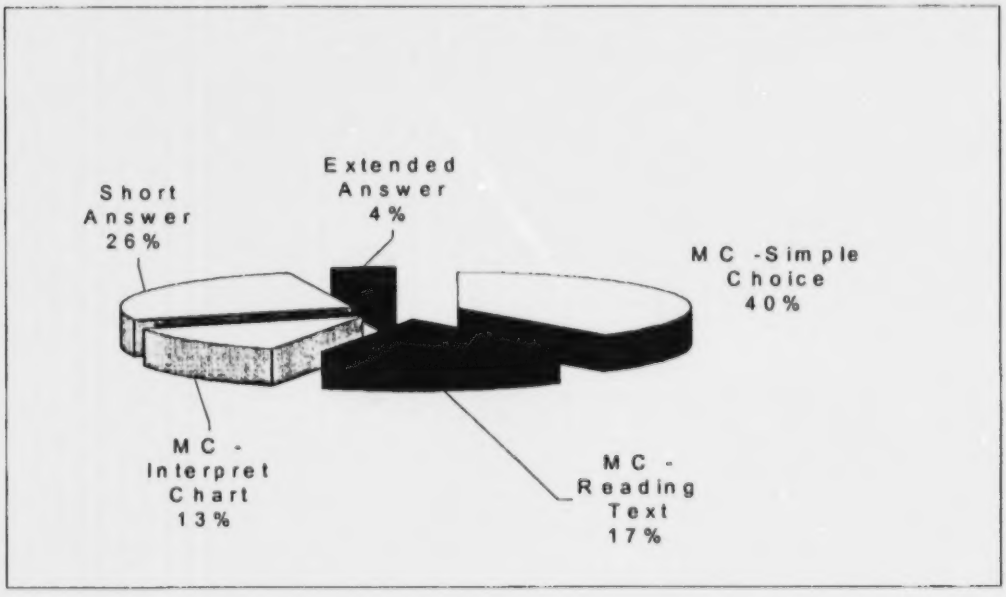
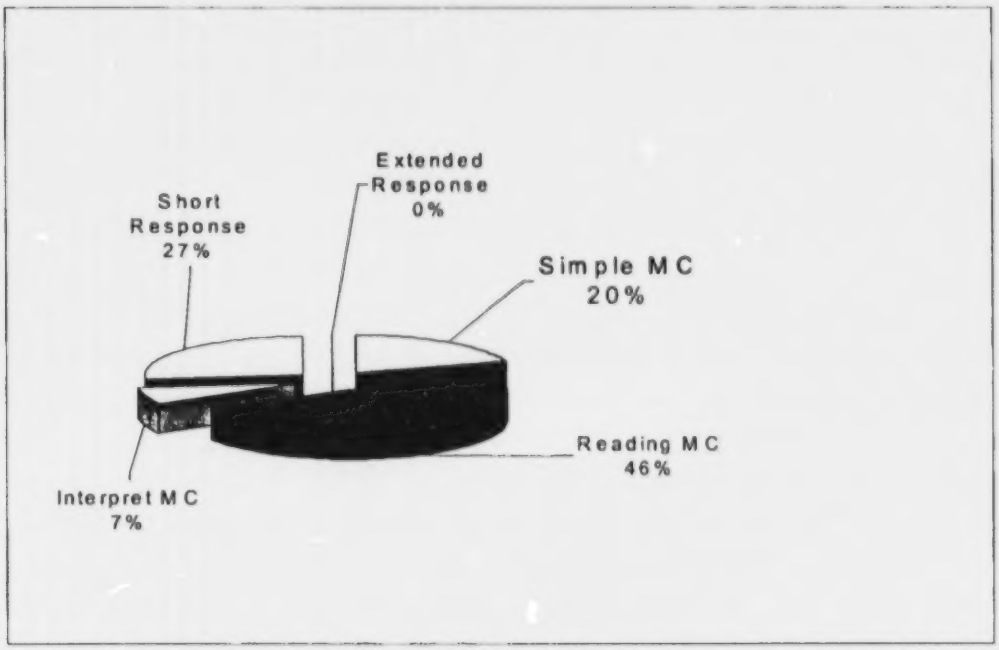


Figure 5 Percentage of Types of Questions for Life Science, 6th Grade OPT

The fourth grade OPT assesses 19 different learning outcomes that are grouped into the same four strands as the sixth grade test. Overall, the fourth

grade OPT shows similar results for analysis of the content with the multiple-choice items dominating the total number of items across 17 of the 19 learning outcomes. Learning Outcomes #3, 4, 7, and 9 favored the multiple-choice items at better than a 4:1 ratio over the short and extended response items. Of the eighty (80) total items for years 2000 and 2001, 34% of them were multiple-choice items that required students to interpret a figure, table, or graph prior to responding. Of the 61 multiple-choice items, almost half (45%) required students to work with a figure, table, or graph. As with the sixth grade test, students must be competent in the use of figures, tables, and graphs to achieve success on the 4th grade OPT.



Analysis of the *life science* strand is depicted in Figure 6. Note that the distribution of items in the *life science* strand deviates from the general overall pattern of distribution with regard to the format and style of the test items. In this

strand, the dominant type of item is clearly multiple-choice items that require students to *read* text followed by *short response* items.

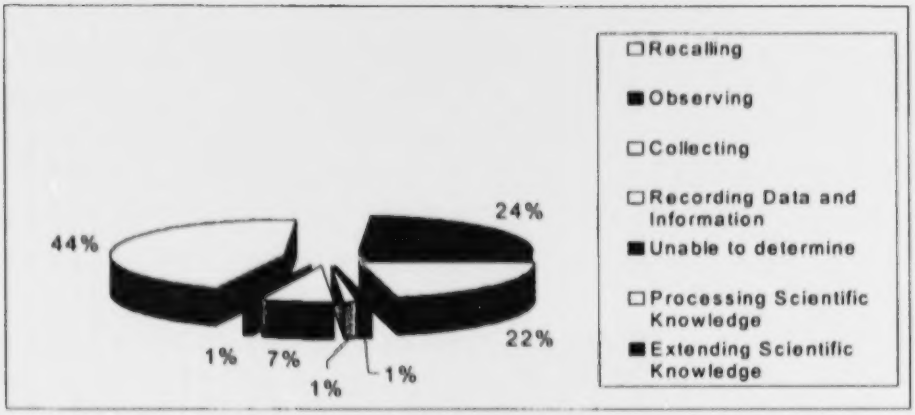
Figure 6 Percentage of Types of Questions for Life Science, 4th Grade OPT

The remaining figures in this section display the results of the kinds of *scientific processes* that the OPT assesses.

The OPT assesses three different scientific processes: *acquiring scientific knowledge*, *processing scientific knowledge*, and *extending scientific knowledge* as depicted in Figure 7. Each of these categories is grouped into two levels or *grains*, which are further divided into additional categories. (Refer back to the methods section for details of these grains.)

Figure 7 Analysis of Process, 6th Grade OPT

The researcher was able to identify 55 of the 92 total *first grain* items for years 2000 and 2001 of the sixth grade OPT. For the second grains, 12 items could not be identified. Only one item could not be identified within the first grain category of *acquiring scientific knowledge*. The researcher was able to identify all of the second grains within the category of *acquiring scientific knowledge*. Overall, the category of *processing* dominated the total number of items. Forty-four percent of the items required students to *process scientific knowledge*. Based on the analysis, the ability to *process scientific knowledge* is an essential skill for succeeding on the 6th grade OPT.



Because the fifth grade teachers in this study assessed primarily in the *Acquiring Knowledge* category, analysis of this category is depicted in Figure 8 for the first set of grains. The four grains in this category consist of *recalling*, *observing*, *collecting*, and *recording data and information*. The researcher was not able to identify less than one percent of the grains in the category of *Acquiring Knowledge*. Results show that, on the 6th grade OPT, approximately 22% of all of the items from years 2000 and 2001 were identified by the researcher as belonging to the first grain category of *recall*. Results also show that there are approximately twice as many items (44%) in the category of *Processing Scientific Knowledge* as there are first level grains of *recalling* in the category of *Acquiring Scientific Knowledge*.

Figure 8 Acquiring Knowledge, Analysis of First Grains, 6th Grade OPT

The *second set of grains* for the process category of *Acquiring Knowledge* consists of three grains - *recall fundamental facts, concepts, or relationships*; *take measurements*; and *read graphs*. The first of these second level grains (*recall fundamental facts, concepts, or relationships*) is depicted as "*recall facts*" in Figure 9. For this grain, the same percentage (22%) of the total number of

items from the tests of years 2000 and 2001 was identified as it was for the first grain of *recall*. Generally, both of these grains from Acquiring Knowledge can be characterized as *recalling information*, which requires a low level cognition. Although *recalling information* is an important part of *Acquiring Knowledge*, notice that of the other two major process categories, *Extending Scientific Knowledge* (24%) is about as important as *recalling information* and *Processing Scientific Knowledge* (44%) is twice as important *recalling information*.

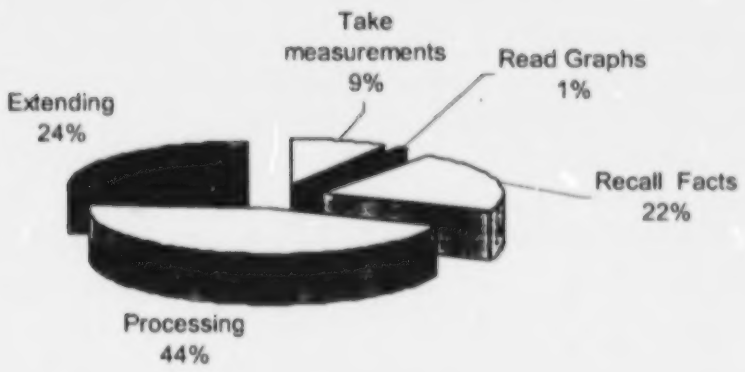


Figure 9 Acquiring Knowledge, Analysis of Second Grains, 6th Grade OPT

The third grade participants in this study also assessed their students primarily in the *Acquiring Knowledge* category. Results of the analysis of this category are shown in Figure 10 for the first set of grains. On the 4th grade OPT,

approximately 25% of all of the items from years 2000 and 2001 were identified by the researcher as belonging to the first grain category of *recall*. No items were found for the first level grain of *collecting* in the category of *Acquiring Knowledge*. Results also show that there are approximately more than 1.6 times as many items in the category of *Processing Scientific Knowledge* as there are first level grains of *recalling* in the category of *Acquiring Scientific Knowledge*.

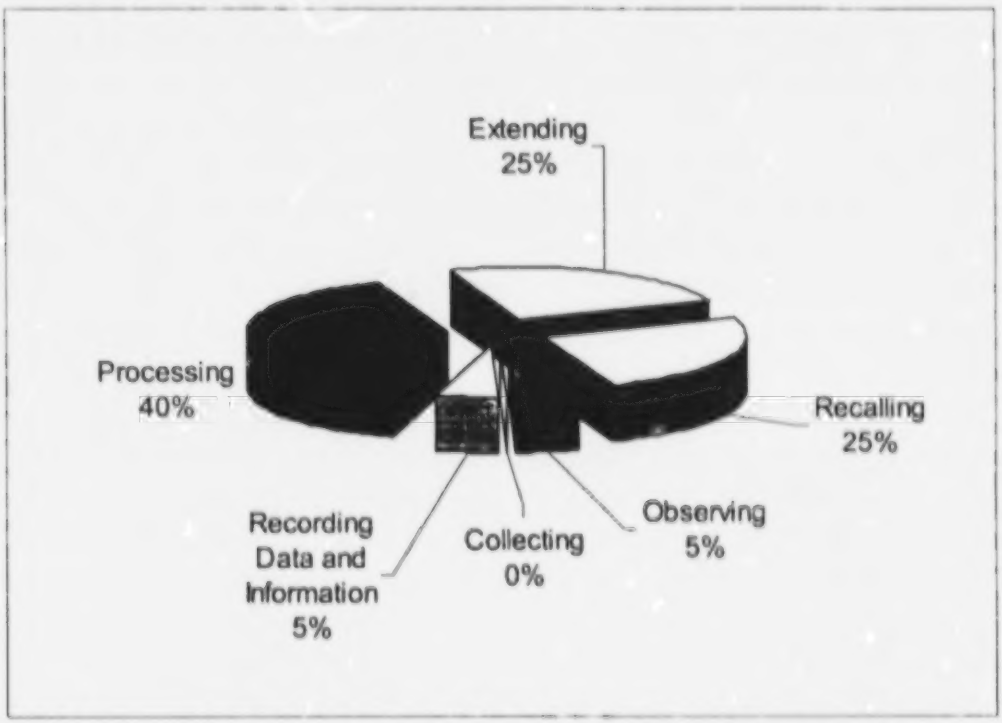


Figure 10 Acquiring Knowledge, Analysis of First Grains, 4th Grade OPT

Because the third grade participants in this study also assessed primarily in the *Acquiring Knowledge* category, analysis of this category is depicted in Figure 11, for the second set of grains. On the 4th grade OPT, approximately

25% of all of the items from years 2000 and 2001 were identified by the researcher as belonging to the second grain category of *recall fundamental facts, concepts, or relationships*. Generally, both of these grains from *Acquiring Knowledge* can be characterized as *recalling information*, which requires a low level of cognition. Although *recalling information* is an important part of *Acquiring Knowledge*, notice that of the other two major process categories, *Extending Scientific Knowledge* (25%) is equally as important as *recalling information* and *Processing Scientific Knowledge* (40%) is much more important.

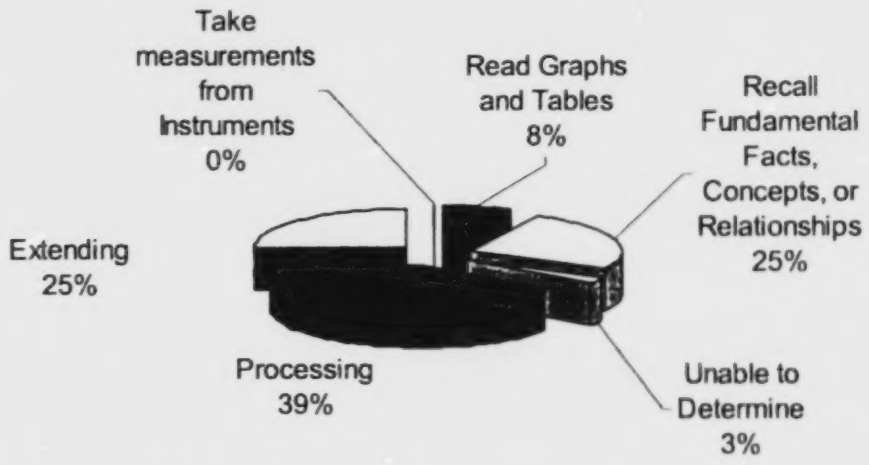

Take measurements from Instruments 0%
Read Graphs and Tables 8%
Recall Fundamental Facts, Concepts, or Relationships 25%
Extending 25%
Processing 39%
Unable to Determine 3%

Figure 11 Acquiring Knowledge, Analysis of Second Grains, 4th Grade OPT

Document Results - Content

Overall, comparison of the OPT data with the formal summative classroom assessment practices of the participants finds little or no similarities. With the exception of one teacher participant, little evidence of alignment with any of the

three categories is found in the participants' document data. The most favorable comparison is with the data from one of the fifth grade teachers. Results of the document data are analyzed according to content, process, and types of items (i.e. style and format). Also reported, is the percentage of teacher-designed items versus publisher-designed items.

Most of the participants' items are not related to any of the 17 (6th grade) and 19 (4th grade) learning outcomes assessed by the OPT. Relatedness was determined using face validity to check whether or not the teachers' items assessed the content described in the learning outcomes found in the Ohio Department of Education's *Guidelines for the Ohio Proficiency Test*. Results of analysis of the content assessed by the teachers' assessment items are displayed in the Expanded Tables of the Appendix. Analyses indicate almost no linkage with the content of the OPT for either of the groups. Table 1 shows that only 3% of the third grade teachers' items can be linked to the learning outcomes assessed by the 4th grade OPT and only 10% of the fifth grade teachers' items can be linked to the learning outcomes assessed by the 6th grade OPT. In fact, two of the fifth grade teachers and one of the third grade teachers have no items that can be linked to any of the learning outcomes assessed by the science OPT. Teachers Z and X showed the best alignment with 62% and 26% of their items relating to the learning outcomes respectively. By itself, this analysis should not be used to reach a definitive finding regarding the participants' alignment with the content of the OPT because the period of investigation only covered the first quarter plus the first two weeks of the second quarter. During this time, the participants were focused mostly on the *life science* strand. However, evidence

provided in the remainder of this section suggests that the same pattern of misalignment exists within the content strand of *life science*.

Table 1

Percentage of Participants' Items Related to the Content of the OPT

Third Grade Teachers	Related	Fifth Grade Teachers	Related
E	2%	Z	62%
D	2%	Y	5%
C	8%	X	26%
B	5%	W	0%
A	0%	V	0%
Third Grade Teachers Combined	3%	Fifth Grade Teachers Combined	10%

Because the teachers mostly focused their instruction on the *life science* strand during the period of this investigation into their assessment practices, that is the strand that the researcher targeted for analysis. The fifth grade results for the life science outcomes are displayed in Expanded Table 7 of the Appendix. Although the fifth grade teachers were primarily focused on *life science*, the results show, with the exception of teacher Z, that most of the assessment items are not linked to any one of the four learning outcomes in this particular strand. For example, in the life science strand, Learning Outcome #16 is not formally assessed by any of the teachers, while only one teacher, using a short response item, assesses learning outcome #17. Analysis of the items related specifically to the sixth grade learning outcomes (#14, #15, #16, #17) for *life science* was performed and is displayed in Table 2.

Table 2

Comparison of the Frequencies of Various Types of Items Related to Learning Outcomes #14, #15, #16, & #17 (Life Science Strand) Between Fifth Grade Teachers and the Sixth Grade OPT

Combined Total of All Fifth Grade Teachers				
TYPES OF ITEMS	OTHER	MC SIMPLE	MC READ	MC INTERPRET	SHORT

			EXTENDED		TOTAL		
Frequency of items	3	17	0	0	15	3	**38**
% of Items	8%	45%	0%	0%	39%	8%	**100%**
6th OPT - years 2000 & 2001							
Frequency of items	0	9	4	3	6	1	**23**
% of Items	0%	39%	17%	13%	26%	4%	**100%**

Although the fifth grade teachers were primarily focused on *life science*, the results displayed in Expanded Table 7 of the Appendix show that, generally, most of the assessment items are not linked to any particular one of the four learning outcomes in this strand. When the fifth grade teachers' items are analyzed as a group, it appears that there is partial alignment with the style and format of the items for learning outcomes #14, #15, #16, and #17 with the *simply constructed multiple-choice*, *short response*, and *extended response* items. However, the data clearly suggest that this alignment is not replicable for each of the separate learning outcomes of the life science strand. Evidence of alignment is only observed when all four learning outcomes are considered as a strand for the combined data of the fifth grade teachers. The results for the categories of *simply constructed multiple-choice*, the *short response*, and *extended response* items are the most comparable. Not certain whether the difference between the fifth grade teachers and sixth grade OPT were significant, the researcher conducted a chi-square test between the combined frequencies of the fifth grade teachers and the OPT's frequency of multiple choice items, short response items, and extended response items for the learning outcomes of the Life Science strand. A hypothesis was developed where H_o: there is a significant difference between the pattern of distribution of frequencies for the two groups. Table 3 shows the actual and expected frequencies for this comparison.

Table 3			
Combined Items for Fifth Grade Teachers Related to OPT Life Science Strand (LO: #14 - #17) Compared to 6th OPT			
Actual Frequency			
	MC Simple MC Interpret		**MC Read**
Fifth Teachers	17	15	3
6th OPT	9	6	1
Expected Frequency			
Fifth Teachers	17.8	14.4	2.7*
6th OPT	8.2	6.6	1.3*

*It should be noted that the third cells for *expected frequencies* are less than 5, which tends to distort the calculations.

A two-tailed chi-square test of significance of the data shows a critical value of 0.861135. The test fails to reject the null hypothesis, $\chi^2(2, n = 35) = 5.99$, $p < .05$, suggesting that the fifth grade teachers' pattern of testing the learning outcomes from the life science strand with these three kinds of items may be similar to that of the OPT. However, the total frequency of the items (38) for eleven weeks of instruction for five teachers would appear to the investigator as inadequate. This number equates to less than eight items per teacher. The very low number of items suggests misalignment between the teacher's fifth grade science curriculum and/or instruction and Ohio's standards as defined by the OPT.

Although the third grade teachers were also primarily focused on *life science*, the results shown in the Expanded Tables of the Appendix, again, show that most of the assessment items are not linked to any one of the four learning outcomes in this particular strand. Only the categories of *simply constructed multiple-choice* and *short response* are the most comparable. In fact, one of the

third grade teachers does not have any items that can be linked to any of the learning outcomes assessed by the 4th grade OPT. Finally, Learning Outcome # 19 is not formally assessed by any of the teachers. Table 4 summarizes the results for the content analysis of the life science strand.

Table 4 Combined Items for Third Grade Teachers Related to OPT Life Science Strand (LO: #16 - #19) Compared to 4th OPT

	Other Extended	MC Simple	Total	MC Read	MC Interpret		Short
Frequency	11	3	0	0	5	0	**19**
%	58%	16%	0%	0%	26%	0%	**100%**
4th OPT - years 2000 & 2001							
Frequency	0	3	7	1	4	0	**15**
%	0%	20%	47%	7%	27%	0%	**100%**

For both groups, it shows an overall lack of alignment between the content of the OPT and the teachers' assessment items. This result suggests one of the following three scenarios - the curriculum is not aligned to Ohio's standards or the instruction is not aligned to the curriculum or the assessment is not aligned to the content of the prior instruction and learning.

Document Data - Process

Results of analysis of the document data for process are displayed in Tables 5 and 6. They show that, for the most part, the tests used in the participants' classroom assessments contain an overabundance of items characterized as low-level in terms of the knowledge products or thinking skills required by the science proficiency tests. Other studies corroborate this finding (Bol & Strage, 1996; Crooks, 1988; Fleming & Chambers, 1983; Gallager, 1991; Haertel, 1986; Madaus, Maxwell West, Harmon, Lomax, & Viator, 1992). Analyses also suggest a high degree of misalignment between the teachers' classroom assessments and the OPT with regard to the kinds of processes that

each assesses.

Table 5 shows that for four of the fifth grade teachers, more than three fourths (¾) of their assessment items are classified as requiring very low-level cognitive skill - *recalling information* as compared to 22% for the proficiency tests. Teacher Z has the most comparable results for *recalling information* in both *grains*.

Table 5 Frequencies of 5th Grade Items for each of the Grains of Acquiring Knowledge

	Acquiring - 1st grain*				Acquiring - 2nd grain			Total for Both		
Fifth Grade Assessments for Recall Facts	Teachers Graphs & Tables		Recall	Observe Recall Facts	Collect	Record Data Total Items		Take Measure-% for Recall	%	
Z	19	0	0	9	0	2	26	42	45%	
	62%									
Y	286	1	0	3	4	0	286	292	98%	
	98%									
X	118	15	0	8	19	4	118	156	76%	
	76%									
W	115	0	0	9	0	9	115	127	91%	
	91%									
V	162	7	0	0	0	7	162	169	96%	
	96%									
OPT * 00 & 01	20	1	1	6	8	1	20	92		
	22%	22%								

* One of the OPT items could not be classified for the first grain with confidence by the investigator

The reliance on low-level items to assess their students is more prevalent with the third grade teachers. Table 6 shows that for four of the five teachers, only one of the grains for one of the teachers has less than ninety (90%) of its assessment items classified as requiring very low-level cognitive skill - *recalling*

info..nation - as compared to twenty-five (25%) for the proficiency tests. The remaining percentages of these low-level questions have a range of 92% to 99% for both grains of *acquiring knowledge*. Teacher A comes closest to the OPT results with forty (40%) of her items classified as *recalling information*.

Table 6 Frequencies of 3rd Grade Items for each of the Grains of Acquiring Knowledge

	Acquiring - 1st grain				**Acquiring - 2nd grain***						
Third Grade Teachers	Graphs & Tables		Recall	Observe		Collect	Re-cord Data		Take	Measure-	
ments for Recall Facts				Recall Facts		Other	Total Items		% for Recall	%	
E	72	9	1	1	0	0	82	1	83	87%	
	99%										
D	145	0	0	0	4	0	145	0	155	94%	
	94%										
C	36	0	1	0	19	0	36	0	39	92%	
	92%										
B	57	0	0	0	0	0	57	0	61	93%	
	93%										
A	19	22	0	0	0	12	19	10	47	40%	
	40%										
OPT * 00 & 01	20	4	0	4	0	6	20	2	80		
	25%	25%									

* Two of the OPT items could not be classified for the second grain with confidence

A comparison of the percentages of the first grain elements for the process category of *Acquiring Scientific Knowledge* with the categories of *Processing* and *Extending* finds significant differences between the fifth grade teachers' percentages and those of the OPT. Table 7 shows that Teacher Z comes closest to aligning with the OPT, but even her use of *recalling* items is

more than double the percentage found for the OPT and her category for *Processing* is only a little better than half of what the OPT shows. Her results for the category of *Extending* are less than a third of what is found from analysis of the OPT items. With the remaining fifth grade participants, the amount of difference is even more profound.

Table 7
First Grains of Acquiring Compared to Processing & Extending for 5th Grade

Fifth Grade Teachers and Information		**Unable to Determine**	**Recalling**	**Observing Processing**	**Collecting Extend- ing**	**Recording Data**	
Z	45%	0%	0%	21%	0%	26%	7%
Y	98%	0%	0%	1%	0%	0%	1%
X	76%	10%	0%	5%	0%	6%	3%
W	91%	0%	0%	7%	0%	2%	0%
V	96%	4%	0%	0%	0%	0%	0%
Combined	89%	3%	0%	4%	0%	3%	1%
OPT	22%	1%	1%	7%	1%	45%	24%

A similar difference is found when the second grains are expressed as percentages. Table 8 also shows heavy reliance on low-level items (*recalling fundamental facts, concepts, or relationships*) for all participants. Clearly, the fifth grade teachers are not using enough items in the categories of *Processing* and *Extending*, indicative of a high level of incongruence between the processes assessed by the OPT and the processes assessed by the fifth grade teachers.

Table 8
Second Grains of Acquiring Compared to Processing & Extending for 5th Grade

Fifth Grade Teachers Graphs and Tables	**Unable to Determine**	**Take measurements from Instruments Recall Fundamental Facts, Concepts, or Relationships**	**Process-ing**	**Extend- ing**		**Read**
Z	0%	5%	62%	0%	26%	7%
Y	1%	0%	98%	0%	0%	1%
X	12%	3%	76%	0%	6%	3%
W	0%	7%	91%	0%	2%	0%
V	0%	4%	96%	0%	0%	0%
Combined	3%	3%	90%	0%	3%	1%

OPT	9%	1%	22%	0%	45%	24%

A comparison of the percentages of the first grain elements for the process of *Acquiring Scientific Knowledge* with the categories of *Processing* and *Extending* also finds misalignment between the third grade teachers' percentages and those of the OPT. Table 9 shows that Teacher A comes closest to aligning with the OPT, but even her use of recalling items is significantly more than the percentage found for the OPT and her category for *Processing* is only a fourth of what the OPT shows. Her result for the category of *Extending* is not significantly different from the others in her group. Overall, the same level of misalignment found with the fifth grade teachers is seen with the third grade teachers.

Table 9

First Grains of Acquiring Compared to Processing & Extending for 3rd Grade

Third Grade and Information	**Teachers**	**Recalling**	**Processing**	**Observing**	**Extending**	**Collecting**	**Recording Data**
E	87%	11%		1%	1%	0%	0%
D	94%	0%		0%	0%	6%	1%
C	92%	0%		0%	0%	5%	3%
B	93%	0%		0%	0%	5%	2%
A	40%	47%		0%	0%	11%	2%
Combined	85%	8%		0%	0%	5%	1%
OPT	25%	5%		0%	5%	40%	25%

The same pattern of misalignment is found when the second grains are expressed as percentages. Table 10 also shows heavy reliance on low-level items (*recalling fundamental facts, concepts, or relationships*) for all participants. Clearly, the participants are not using enough items in the categories of Processing and Extending.

Table 10

Second Grain of Acquiring Compared to Processing & Extending for 3rd Grade						
Third Grade Teachers		**Take Measurements from Instruments**				**Read**
Graphs and Tables		**Recall Fundamental Facts, Concepts, or Relationships**				
	Unable to Determine		**Process- ing**	**Extend- ing**		
E	0%	0%	99%	1%	0%	0%
D	0%	0%	94%	0%	6%	1%
C	0%	0%	92%	0%	5%	3%
B	0%	0%	93%	0%	5%	2%
A	0%	26%	40%	21%	11%	2%
Combined	0%	3%	88%	3%	5%	1%
OPT	0%	8%	25%	3%	40%	25%

Document Results - use of black-line masters, teacher designed tests, and rubrics

It is not uncommon for the publishers of textbooks to include a package of assessment items and tests or "*black-line masters*" that can be copied by the teachers and used to assess their students' learning. Document data provide evidence that the teachers in this study rely heavily on the use of publisher's black-line masters to assess student learning. Table 11 compares the use of black-line masters to teacher-designed tests. During the period of this investigation, the teachers used black-line masters for 73% of their formal assessments. Only 23% of their assessments were teacher-designed. These results differ from the teacher's perceptions of their practices described later in this study. This discovery also differs from the findings of previous researchers (Haynie, 1983, 1991, 1992; Herman & Dorr-Bremme, 1982; Mehrens, 1987; Mehrens & Lehmann, 1987; Newman & Stallings, 1982) who found most teacher assessments were done with teacher-made tests.

Table 11 compares the teachers perceptions gleaned from the questionnaire data with the actual practices of the teachers as determined by the

document data. In addition to the total number of tests, quizzes, and projects that were used by the teacher to grade the students, it shows the number of rubrics submitted and the percentage of teacher-designed assessments versus the percentage of the textbook publishers' assessments used by the teachers. Of significance is the result that shows seventy-three percent (73%) of all of the teacher's formal summative assessments were derived from the package of assessments provided by the adopted textbook series.

Table 11

Comparison of Document Data with Questionnaire Data for Various Kinds of Assessments

	Document Data (Actual use)				Questionnaire Data (Perceived use)		
	Number of Total Assessment Documents				**Number of Rubrics Provided**		
	Teacher Designed Assess-ments		**Publisher Black-line Assess-ments**				
	Teacher Designed Assessments		**Publisher's Black-line Assess-ments**				
	Other						
E	11		36%	64%	20%	20%	60%
D	14	1	29%	71%	5%	60%	35%
C	5	1	60%	40%	10%	55%	35%
B	6		0%	100%	40%	20%	40%
A	7		0%	100%	40%	20%	40%
Z	4	2	100%	0%	20%	50%	30%
Y	17		41%	59%	20%	30%	50%
X	19	4	21%	79%	50%	0%	50%
W	10		0%	100%	35%	35%	30%
V	14		21%	79%	25%	20%	55%
Totals	**107**	**8**	**27%**	**73%**			

The results from Table 11 also show high levels of incongruence between

the actual use and the perceived use for several of the teachers such as Teachers B, Teacher A, and Teacher W. The areas of greatest difference are highlighted. Curiously, Teacher Z, who demonstrated the most alignment throughout this study, perceived more reliance on the publisher's package of assessments than what she used during the period of this study.

Finally, the teachers were also asked to submit any rubrics if they were used to assess the responses or projects. Only four teachers submitted rubrics. However, of the eight rubrics that were submitted, all of them were fairly vague and lacking any real descriptive criteria for assessing the students' work. For example, most of the rubrics relied heavily on "criteria" such as these four items - "*organized*", "*all necessary topic areas on poster*", "*neatness*", and "*teamwork*". For a seed collection project, another rubric offered, "*one point for every seed collected up to 10 points*", "*five points for student's name and section*", and "*five points for neatness*". Only one of the criteria with this rubric addressed science content or processes and its level of cognition was very low because it did not contain a requirement for accuracy - "*one point for each seed that was labeled*".

Document Results - format and style

Table 12 shows that a significant number of the teachers' assessment items fall into the category of *other* (i.e. they are not in the style or format of the proficiency test). Analysis of the data found that 71% of all of the participants' items are not aligned to the style and format found on the OPT. These items generally fall into three categories - true/false, completion, and matching. The remaining items are mostly multiple-choice.

Table 12
Percentage of All Items Consisting of Types other Than Those Found on the

OPT

Third Grade Teachers		Fifth Grade Teachers	
E	87%	Z	12%
D	66%	Y	69%
C	72%	X	71%
B	48%	W	85%
A	68%	V	85%
Combined	69%	**Combined**	72%
Both groups combined = 71%			

Analysis of the format and style of the third grade teachers' multiple-choice items revealed that 100% of them are *simple*. For the purpose of this investigation, *simple* is defined as items that *use one sentence to ask the question* or *one sentence could have been used to ask the question*. Table 13 compares the data for third grade teachers with the OPT analysis. The results suggest that the participants' multiple-choice items are not aligned with the style of the multiple-choice items on the OPT. The analysis shows that simply constructed multiple-choice items are overly emphasized in their assessments while not nearly enough emphasis is placed on the use of multiple-choice items requiring students to read several sentences and/or interpret graphs or table or figures.

Table 13
Frequency of Various Types of Multiple-Choice Items

Third Grade Teachers	Simple	Read	Interpret
E	11	0	0
D	40	0	0
C	2	0	0
B	26	0	0
A	10	0	0
4th OPT for years 00 & 01	**12**	**21**	**27**

Table 14 Comparison of OPT's Frequencies of Multiple-choice Items with

Frequencies of Teacher Z's Multiple-Choice Items

Fifth Grade Teachers	**Simple**	**Read**	**Interpret**

For the fifth grade teachers, the same finding was reached except for one teacher (*Teacher Z*). Table 14 illustrates how the results of the participants contrast with the results from the analysis of the sixth grade science OPT.

Z	9	5	3
Y	52	0	0
X	25	0	0
W	25	0	0
V	13	0	0
6th OPT for years 00 & 01	25	11	32

Teacher Z's 17 multiple-choice items rank fifth for all participants with the number of items ranging from 52 to 2. The mean for the frequency of multiple-choice items of all teachers is 22.1, the mode is 25, and the median is 27. Since *Teacher Z* was the only participant with other types of multiple-choice items, her data is compared to the analysis of the OPT. Her results are shown together with those of the OPT in Table 15.

Table 15 Comparison of OPT's Multiple-Choice Items with Teacher Z's Multiple-Choice Items

OPT Sixth Grade Science (Years 2000 & 2001)			
68 Multiple-Choice Items			
Statistic	**Simple**	**Read**	**Interpret**
Frequency	25	11	32
% of MC Items	37%	16%	47%
% of All 92 Items	28%	12%	35%
Teacher Z (First Eleven Weeks of School Year 2001-2002)			
17 Multiple-Choice Items			
Statistic	**Simple**	**Read**	**Interpret**
Frequency	9	5	3
% of MC Items	53%	29%	18%
% of All 42 Items	21%	12%	7%

As part of the analysis, a chi-square test was performed between Teacher Z and the OPT's frequency of various types of multiple choice items where H_o: there is a significant difference between the pattern of distribution of frequencies for the two groups. Table 16 shows the actual and expected frequencies for this comparison.

Table 16

Comparison of Teacher Z and OPT for Types of Multiple-Choice Items

Actual Frequencies

	Simple	**Read**	**Interpret**	
6th OPT	25	11	32	68
Teacher Z	9	5	3	17
	34	16	35	85

Table 16 *Continued*

Expected Frequencies

	Simple	**Read**	**Interpret**	
6th OPT	27.2	12.8	28	68
Teacher Z	6.8	3.2*	7	17
	34	16	35	85

*It should be noted that the second cell for Teacher Z's *expected frequencies* is less than 5, which tends to distort the calculations.

A two-tailed chi-square test of significance of the data shows a critical value of 0.081575. The test fails to reject the null hypothesis, $\chi^2(2, n = 17) = 5.99$, $p < .05$, suggesting that teacher Z's pattern of testing with multiple-choice items may be similar to that of the OPT. This finding is supported by her interview comments and her questionnaire disclosures discussed in the remaining sections.

Teacher Questionnaire Results

The teacher questionnaire consisted of two parts. Part I required the teachers to indicate which one of the five rubric descriptors most applied to each of the learning outcomes assessed by the OPT. The rubric descriptors were assigned a predetermined value by the researcher after the questionnaires were collected - "*not relevant to this grade level or not addressed by the schools' printed or taught curriculum*" was assigned zero points; "*minimal instruction with no formal assessment*" was assigned one point; "*instruction with informal assessment using checklists or teacher observation*" was assigned two points; "*instruction with formal assessment using publisher's black-line masters or tests*" was assigned three points; "*instruction with formal assessment using teacher-designed tests with T-F, matching, simple multiple choice*" was assigned four points; "*instruction with formal assessment using teacher-designed tests or projects aligned to the style/format of the OPT*" was assigned five points. For the purpose of this study, formal assessment was defined to the participants as a *paper and pencil test* or a *project with a rubric similar to the rubrics used to assess extended and short responses on the OPT whose results are counted as part of the students' summative evaluations.*

Teachers' responses to Part I of the questionnaire suggest that, overall, they believe that the content being assessed by their questions is aligned to the Learning Outcomes being assessed by the OPT. Only two teachers reported not assessing some of the learning outcomes at least informally. Teacher B (3rd grade) reported a rubric score of "0" for one of the learning outcomes and Teacher X (5th grade) reported a rubric score of "0" for four of the learning outcomes. However, for all teachers in both groups, 92 outcomes out of a

possible 180 were reported as not being *formally* assessed. While the teachers believe that they are assessing the same content as is tested by the OPT, they report formally assessing only half of the outcomes. Their responses, shown in Expanded Table 9 of the Appendix, indicate that the remaining outcomes are assessed using informal practices such as checklists and teacher observation.

For the third grade teachers, averaging their assigned rubric scores for each of the learning outcomes yields a range of means from 1.9 to 2.9. A score of less than a "3" on the questionnaire infers that the specific learning outcome was not formally assessed and a score of less than "5" infers that assessment of the specific learning outcome was not aligned to the style and format of the OPT. The mean of the combined means for all third grade rubric scores is 2.4. This result suggests that the third grade teachers perceive that they are essentially assessing the learning outcomes with a balance of formal and informal assessments but that they seldom assess the learning outcomes with items aligned to the style and format of the OPT.

In contrast, the average rubric scores for the fifth grade teachers yield a range of means from 3.1 to 4.3. The combined means for all fifth grade rubric scores is 3.6 indicating that the teachers perceive that they are formally assessing their students with a mixture of black-line masters and teacher-designed tests or projects. Clearly, the fifth grade teachers perceive that they are assessing their science students at a higher level than the third grade teachers. This result is not congruent with the results of the document data, which show a heavy reliance on black-line masters and almost no use of teacher-designed assessments.

To determine if the teachers perceive any difference in their level of assessment of the life science strand as compared to the other strands, the researcher compared the scores from the four learning outcomes in this strand with the averages of the remaining strands. The mean for the third grade teachers perceived level of assessments for the life science strand is 2.65 as compared to 2.35 for the remaining strands. The mean for the fifth grade teachers perceived level of assessments for the life science strand is 4.3 as compared to 3.4 for the remaining strands. These results suggest that both groups perceive that they are assessing the life science strand with items that are more closely aligned to the OPT than the other strands.

Analysis of fifth grade data from Part I of the questionnaire reveals that the average rubric scores for the learning outcomes of the *Life Science* strand has a range of 3.5 to 5 with a mean of 4.3 and a standard deviation of 1.02. This result compares to a range of 2.5 to 4.1 with a mean of 3.4 and a standard deviation of 1.55 for the remaining strands. These results would suggest that the fifth grade teachers perceive that during the period of this study, they predominantly assessed their students with self-designed tests and that they perceive that they will rely more heavily on black-line masters for assessing the other strands. However, analysis of Part II of the questionnaire contradicts the result. The result shows that only one teacher believed they used self-designed assessments at least half of the time. One finding is clear - both sources of questionnaire data are congruent with the results of the document data which suggests that the teachers do not use assessments that are aligned to the style and format of the OPT.

Data from the teachers' responses for the learning outcomes specific to the life science strand was specifically examined for triangulation with the findings of the document data. Tables 17 and 18 show the frequency of items found in the document data assessing each of the life science learning outcomes. An asterisk next to the frequency of items indicates that the teacher perceives that they formally assessed the outcomes. Recall that all formal assessments were collected from the teachers during the period of instruction that focused on the life science strand. The outcomes where no items were found that are also marked with an asterisk are considered as examples of extreme incongruence.

Table 17 Triangulation of Document Data and Questionnaire Data

Third Grade Teachers	LO #16	LO #17	LO #18	LO #19
E	2*	0*	9	0
D	2	1	1	0
C	2	1	0	0
B	0*	1*	0	0*
A	0*	0*	0	0*

* Teacher perceived that they were formally assessing the Learning Outcome

Table 18 shows that based on the questionnaire data, only Teacher Y does not perceive assessing all of the learning outcomes from the life science strand of the sixth grade science OPT. Learning Outcome #14 is the only outcome not assessed by Teacher Y.

Table 18
Triangulation of Document Data and Questionnaire Data

Fifth Grade Teachers	LO #14	LO #15	LO #16	LO #17
Z	16*	7*	0*	1*
Y	10	1*	0*	0*
X	3*	0*	0*	0*
W	0*	0*	0*	0*
V	0*	0*	0*	0*

* Teacher perceived that they were formally assessing the Learning Outcome

To calculate the *rate of agreement*, the researcher tallied these extreme examples of incongruence and divided by the total number of possible examples (20). Results for the third grade teachers show a *rate of agreement* of 70% while the fifth grade teachers' *rate of agreement* is 30%. The results suggest that the perceived level of assessment practices for the third grade teachers is aligned more with their documented practices than the fifth grade teachers.

Based on these results, the fifth grade teachers perceive that they are formally assessing their students more frequently than the third grade teachers. However, both sets of data reveal very low frequencies for an entire quarter of formal assessments. Only four examples of frequencies higher than three are found in the data. These data would suggest that the teachers' perceptions generally are not aligned with their practices.

With regard to whether the teachers' tests are aligned to the textbook or the OPT outcomes, analysis of the mean scores indicates that the fifth grade teachers believe they assess most of the learning outcomes with a combination of teacher-designed tests and black-line masters. Table 19 shows that the mean for the entire fifth grade teachers' reported rubric scores for all seventeen of the learning outcomes is 3.6 with a standard deviation of 1.48. This mean score indicates that the fifth grade teachers perceive that they are essentially using a balance of teacher-designed tests and publisher's black-line tests to formally assess their students. However, the mean also indicates that the teachers' believe that their assessments are not as aligned to the style and format of the proficiency tests as their interview data suggests.

Table 19 shows that the mean for all third grade teachers' reported rubric

scores for the nineteen learning outcomes is 2.4 with a standard deviation of 0.48. This mean score indicates that the third grade teachers perceive that they are essentially using a balance of publisher's black-line tests and informal assessment techniques such as checklists and teacher observation to assess their students. As with the fifth grade teachers, the mean also indicates that the teachers' believe that their assessments are not as aligned to the style and format of the proficiency tests as their interview data suggests.

Table 19

Rubric Scores from Part 1 of Questionnaire

	Combined Means	Standard Deviation
Fifth Grade Teachers	3.6	1.48
Third Grade Teachers	2.4	0.48

Part II of the questionnaire asked the participants to indicate what percentage of their students' grades were comprised of the following: *teacher-designed tests, publisher's black-line masters, student participation, oral exams, group projects, individual projects, portfolios, homework,* and *other.* For the fifth grade teachers, the responses to part II of the questionnaire are congruent with the responses from part I in that they essentially show a balance between the top two types of assessments - teacher-designed and publisher's tests. An average of the percentages was calculated and it shows the fifth grade teachers perceived that, during the period of the study, they used teacher-designed tests for 30% of the student's grades, followed closely by the publisher's tests at 27%. The balance of their combined evaluations consisted of using homework (11%), student participation (10%), with the remainder consisting of oral exams, group

projects, individual projects, portfolios, and other. However, a result that clearly supports the finding from analysis of the document data is that, with one exception (Teacher X), the fifth grade teachers perceive that they developed their own assessments less than half of the time.

For the third grade teachers, the responses to Part II of the questionnaire are not as congruent with their responses from Part I. The averages of the results show the top choice for their summative assessments to be *publisher-designed tests* (35%) followed by *teacher-designed tests* (23%). The category of "*other*", which would include informal techniques such as observation and checklists, was not selected by any of the third grade teachers.

While it would appear that the third grade teachers' perceptions were more congruent with their practices than the fifth grade teachers', overall, the teachers' perceived use of assessment techniques is not congruent with the researcher's analysis of their collected documents. Based on analysis of the interview data described in the next section, this finding suggests that the teachers in this study may not have had enough of an understanding of the OPT and their own assessment strategies to develop a valid alignment between their assessment practices and the assessment format of the OPT or offer an informed description of their personal practices.

Teacher Interview Results

This section describes teacher responses about their assessment practices obtained from individual interviews. Table 20 shows the eight themes that emerged from analysis of the transcribed data.

Table 20
Emergent Themes from Analysis of Transcribed Data using Folio Views

#1	Participants' beliefs regarding alignment between their formal assessments and the OPT
#2	Amount of reliance on publishers' black line masters
#3	Participants' beliefs regarding satisfactory alignment between the publisher's tests and the OPT
#4	Participants' beliefs regarding alignment between their formal assessments and their instructional objectives
#5	Participants' beliefs regarding alignment between their curriculum and Ohio's Learner Outcomes for the OPT
#6	Belief that use of format and style with regular assessments will help students succeed on the OPT
#7	Feelings regarding OPT and whether it is assessing the right things
#8	Probing reveals participant is misinformed or confused about the style, format, processes, and content of the OPT

Table 21 presents the pattern of responses according to analysis of the eight emergent themes. The numbers in the columns represent the themes from the table above. The last row expresses the percentage of agreement found among all participants with regard to the majority response for each theme. The strongest levels of agreement emerged with the themes numbered as 1, 4, and 6. These themes are discussed first.

Table 21
Pattern of Responses According to Analysis of Major Emergent Themes

Teacher	#1	#2	#3	#4	#5	#6	#7	#8
A	Aligned	Most of the time		Yes	Yes	Mostly	Yes	Negative
	Yes							
B	Aligned	Most of the time		Yes	Yes	Misinformed or Confused or		
Not Sure	Yes	Negative	Yes					
C	Aligned	Most of the time		Yes	Yes	Mostly	Yes	Low level
of awareness		Low level of awareness						
D	Aligned	Most of the time		Yes	Yes	Misinformed or Confused or		
Not Sure	Yes	Negative	Yes					
E	Aligned	Most of the time		Yes	Yes	Mostly	Yes	Negative
	Yes							
V	Aligned	Most of the time		Yes	Yes	Misinformed or Confused or		
Not Sure	Yes	Negative	Yes					
W	Aligned	Not much	No	Yes	Mostly	Yes	Negative	Yes

X	Aligned	Most of the time	No	Yes	Process	not aligned		
	Yes	Can live with it	Yes					
Y	Aligned	Most of the time	No	Yes	Misinformed or Confused or			
Not Sure	Yes	Neutral	Yes					
Z	Aligned	No	No	Yes	Mostly	Yes	Positive	No
Level of agree-ment with majority					Aligned 100%		Most of the	Time
(80%)	Yes (60%)		Yes (100%)		Mostly (50%)	Yes (100%)		Negative
(60%)	Yes (80%)							

According to the transcript data both levels of participants in this study perceived that their formal assessments are aligned with all aspects of the OPT. This result differs dramatically with the document data and, to a lesser degree, with analysis of the questionnaire data. This difference suggests that there is significant incongruence between the expressed perceptions of the participants and their actual assessment practices.

All of the participants expressed the belief that their formal assessments and their instructional objectives are aligned. While this belief was shared by all of the participants, analysis of the transcript data suggested that, in some cases, their instructional objectives might not be aligned with the district's printed curriculum.

Elsewhere in the transcript data, evidence was found that the district's adopted curriculum may not be aligned with the objectives assessed by the OPT. For example, teacher Y (fifth grade) explained that when the district developed the curriculum, "*we looked at the fifth grade off year test and the fifth grade Ohio Model outcomes.*" A third grade teacher (B) responded with a similar statement -

> "*I am assuming that when the course of study was formulated that they looked at both the third grade off year and fourth grade proficiency test. I am sure that they did because our curriculum correlates with a lot of what is on the off-year test.*"

"*Off-year test*" is a term coined by publishers wishing to sell standardized

tests to school districts in Ohio for assessing the progress of students in those grade levels that are not administered the state's official proficiency test. The off-year tests are standardized tests purchased by the school district from private publishing companies that are not associated with the Ohio Department of Education. During the course of this study, the researcher spoke with consultants from the Ohio Department of Education who were unwilling to say that these "*off-year tests*" are assessing the same learning outcomes as the OPT. Compounding the issue is the belief by some teachers that Ohio actually produces these off-year tests. A third grade teacher (D) noted that when the district developed its science curriculum, it "*used Ohio's off year proficiency tests.*" In addition, phone interviews with consultants from the Ohio Department of Education suggest that the Ohio Model Curriculum that several of the teachers indicate was used during the curriculum development process was never directly aligned with the content assessed by the OPT. In fact, one consultant noted that different committees were responsible for the development of the Ohio Model Curriculum and the development of the proficiency test items. According to the consultant, these committees never coordinated their efforts with each other.

Finally, all of the participants in this study stated the belief that using the same format and style with their own assessments as found with the OPT would enhance the performance of their students on the fourth or sixth grade Ohio Proficiency Tests. A third grade teacher provided a typical response,

> "*Yes, I think it can. It can be enhanced by using the same style tests. If they have seen that format and have experience with that format in other testing situations, maybe they will be more relaxed and more comfortable when taking the proficiency test. That is one reason why, in third grade, we used to give the off year tests because they need to see that style. Now we* (the district's leaders) *are talking about using normed tests*

(IOWA), which is taking us away from that format again. We are going to have to use it in our regular assessment if we are not going to test that way. It would probably be better if they experienced it (the same format) throughout the year on their tests, because just like riding a bike, the more you practice, the better you get. I think once a year doesn't do it."

Despite their agreement that using the same format as the OPT with their classroom assessments can enhance their students' performance on the OPT, only one participant revealed an accurate understanding of the style and format of the proficiency test items. Upon deeper probing during the interviews, most of the participants admitted basing their understanding of the OPT items on second hand information or the practice test, which is offered only as a scaled-down version of the actual OPT.

One participant showed a very low level of awareness for the style and format of the OPT. For example, when asked about her knowledge of the OPT, she gave these somewhat disconnected, irrelevant responses.

> "*They both use prior knowledge and thinking that they have learned. it's hard to explain. Obviously my tests are going to be over what they have already learned. I think I try to cover what is going to be on the OPT... Somewhat. I think there are some that when we do the OPT, I haven't taught those things because its things that they may have learned later. I think they are aligned about 80%.*" Interviewer: Do you pattern your written tests after the OPT in terms of style/format? "*Yes. I use true and false which I know they use... I think I am trying to use the similar format and style, but I think, you know the questions - I guess I am contradicting myself in a way, because.... You know we do a lot of that bubbling to prepare for it and like the short responses and things like that...* [Name of another teacher] *and I went over these* [the questions from the researcher's interview guide] *together so you are probably going to hear the same things. I think I covered what I needed to say. I hope that is what you wanted."*

Immediately after the session, the following observation was recorded in the interviewer's field notes. *The participant was extremely nervous and seemed determined to give me what she thought I wanted to hear. I doubt the validity of*

some of her answers. She admitted that she was nervous and wanted to give me the answer I was looking for. She was concerned that her answers match her colleague's answers. The researcher concludes that the participant's anxiety rose as a result of her inability to offer a reasonable understanding of the OPT upon deeper probing during the interview. While the validity of her answers is a concern, it is noteworthy that with so much attention given to preparing for the interview, the participant admittedly was not prepared to discuss the OPT.

How well some of the teachers understood the proficiency test was brought into serious question during the interview phase of the study. Probing by the researcher during the interview elicited comments that indicated less than an adequate understanding of the OPT than what would be necessary to effectively align their classroom assessments to its processes that are assessed and style/format of the items. As an example, analysis of the transcript data finds five of the participants believing that the OPT contains true/false questions. The OPT does not use true/false items.

Other evidence found in the transcript data suggests that three other participants also lack an effective understanding of the OPT. For example, one teacher admitted that she had only second hand knowledge of the OPT. "*Last year all of the teachers were given the fourth grade tests, but I was absent that day, so I have not seen it. I can only compare by what I have heard from fourth grade teachers...and how they* (the Ohio Department of Education) *are trying to trick the kids.*" When responding to whether or not she remembered anything in particular about the multiple-choice questions, another teacher made this observation; "*Not exactly, but I think there were quite a few choices including 'all*

of the above and none of the above". [The researcher found no such choices on any of the versions of the OPT that was analyzed.] Another teacher responded to the question about what she knew about the OPT with, "*I know very little* (about the tests) *other than just general information that we have been given and the practice test. As far as discussing the tests, I have not done that at all with the fourth or sixth grade teachers… I don't know everything about the proficiency test. I might learn more in the next couple of years.*" Contrary to what the researcher's analysis of the OPT found, this particular teacher believed that "*the main focus of the proficiency test is short answer and extended response. There are more items that are short answer and extended answer than of the others.*"

As reported in the results of the document data, the teachers in this study rely heavily on packaged tests that accompany their adopted textbook. Analysis of the interview data found supporting evidence with 60% of the participants perceiving that the textbook company's assessment package is aligned with the OPT. For example, teacher E was very satisfied with the publisher's tests.

> "*I am pleased with their* (the publisher's) *tests for the last two years. I have not noticed anything missing. Quite often they align but you can't run all the tests or you would just be testing them constantly, they would never get to do projects. There is a whole booklet of assessment and we pick and choose. I don't use the whole test sometimes. I use a part of the test - a page or two.*" Teacher V: "*I try to use the textbook's tests because they seem to have the same format that shows up on the Ohio tests. From what I have been able to gather about them* (the Publishers tests and the OPT) *I think they are getting closer to the proficiency style. The new series in science is closer. If there is a weakness, it is with their extended response questions.* "

While 80% of the participants perceived that their assessments are aligned with those of the OPT, some at least acknowledged that the publisher's tests on which they so heavily rely are designed differently than the OPT,

specifically noting alignment problems with short answer items, process, and content. Teacher Y noted that the publisher's tests…

> "*…don't have as many extended response items as the* (OPT) *test. The short answer is more fill in the blank and not really a short answer. I don't think they have as many (short answer items). It is more matching, fill in the blank, and multiple choices.*"

Teacher X noted problems with alignment of the processes of science,

> "*I don't think the publisher's tests are aligned with the proficiency because they are more aligned to content whereas the proficiency is more reading and processing.*"

Teacher W offered the most in-depth understanding of the problem of misalignment between the textbook and its accompanying package of tests and the OPT.

> "*There is not a text out there that is aligned with our proficiency. It makes it really difficult for teachers. It is a totally different way of teaching… The testing program* (from the publisher) *that we use does not correlate with the proficiency test. That is the biggest problem. I think we are all having to become proficiency oriented. Most of the testing that we do now is multiple-choice, maybe some short answer…It seems like the testing program that comes with our curriculum is not aligned with the proficiency test…The proficiency test is more process oriented and the textbooks are more fact based - absolutely. It seems like even the newer textbooks, like math for example… we can't even find a math program that is proficiency based. That is the biggest problem with our students - they are so used to having facts fed to them and they can regurgitate those facts, but the minute you make them think - that is our big problem… but here again is the problem. Unless the teacher totally changes the* (publisher's) *testing program, it's not going to happen. The materials are not aligned with the proficiency at all. And so that is why a lot days my kids don't think we have done science because we haven't been in the book. That is when I think that as teachers what we really need to do is really have the goals and objectives both district-wide committed to memory. You have got to know where you are going. To be very honest, we have always had those graded course of study in our little red binders and up until the last few years - yea -- I knew where mine was, but it wasn't used a great deal because we had to turn in papers that we were getting through the book. The curriculum was really the book. The curriculum and the book are now two different things.*"

One teacher (X) recognized that process is an important part of the OPT but she readily admitted that she lacked a thorough understanding of the OPT and the curriculum.

> *"I do more with process than it talks about in our course of study…Process is how I make my link to the proficiency test because I don't feel that our course of study does that. I feel like the kids could not understand anything about the content, but if I taught them how to read and how to expand their thinking, they could answer anything on the proficiency test whether they've had the content or not… I am still learning the course of study and the proficiency test. I am not sure about the content that I am to cover. I don't feel that I know enough to say how well aligned the curriculum is to the proficiency."*

In addition to the main themes that were found in the transcript data, there were several others that may provide insight into why the teachers lack a better understanding of the OPT. For example, many of the teachers expressed strong negative feelings about the OPT such as resentment for the perceived increase in accountability and the call for changes in their instructional and/or assessment strategies. Many, such as teacher D, expressed disagreement with the concept of the design or purpose of the test.

> *"Just the way the questions are worded, it seems to be confusing to some of them. It's not written in a way that is understandable to children of that age. It seems tricky. I do think that it talks about processes a lot, but when they are worded in such a way that is confusing. I know we took the test in our building and I remember on the science, we would argue over the answers to the questions because depending on how you read the question and how you personally determined what it was asking, you could come up with a different answer. When we would find out the real answer, a lot of us were shocked."*

Others, such as Teacher E, spoke of perceived loss of academic freedom. *"Well I have 27 years of experience and I think the OPT has threatened teaching because I don't feel the freedom to go off on a tangent. I don't feel the freedom to go into an interest area.* [She references a recent lesson that deviates from

the written curriculum.] *I feel personally, 'Oh my God, I varied from the curriculum, I'll go to teacher prison or something.'* "

Still others, such as teacher X, spoke of an increase of stress for teachers and students related to the OPT. "*I think that it's* (the OPT) *stressful on everybody. I feel stress that I have to get my kids to pass when there are so many underlying circumstances and I think the kids are stressed because it is so involved.* "

Teacher W was the most open of all the participants regarding the question about the OPT's fairness and appropriateness.

> "*Honestly, I'm not sure what their (the OPT's) objectives really are about. And it seems like just when we start to understand it, they (ODE) change it. I really have a lot of issues with the test because it is very confusing, very tricky. Using very uninteresting stories, using words that are unfamiliar to kids such as unusual names…I think it is an extremely long, difficult, draining test and they are children and they don't really see the significance of it. What is the point really? The outcomes are not so clear and it seems the questions are more about thinking. It seems like we can get away from memorizing facts, but be able be able to read and apply it. But here again it is very difficult and not fair. It is not fair at all because you've got this huge paragraph that most of them can't read. It's too tough. It's very unrealistic. I see some advantages to it that these kids need to know how to think. That is what gets us through as adults - we don't remember everything. I think we are making progress, but it is a tremendous amount of pressure. Teaching is not fun anymore. It is never going to be the children that drive me out, but it is going to be the pressure of trying to exceed up here when you have this down here. I really object to the fact that they say there will come a time when my name will be listed with these children who have or have not passed these tests. We end up teaching to that test and I have never felt that is right. It has taken me a long time to get my mind geared to it. Maybe, for the younger teachers it will be easier, but that basically is what teaching is now - teaching to that test - and just the minute we think we know what we are doing - it's changed - and the standards are changed. Let's show the results, but let's also show the kids IQ. Let's show the whole picture. These proficiency tests do not show the whole picture. It does not show the whole picture of what I'm doing, what these children are doing. Its - 'can they take this test made by these people in this ivory tower.' They're sitting there with their middle to upper class background devising this test for their children or better. I would love for someone to tell me what am I*

doing wrong - what can I do better for these kids. Another thing those proficiency tests don't show is the fact that maybe my kids won't pass that test, but maybe I've made some progress with them - that doesn't show that - its either a pass or fail. Lets also publish where they were at the beginning of fifth grade and where they are at the end of fifth grade…I have made a fraction of a difference, we've made some progress. But none of this testing ever shows that and it is so unfair for Columbus or our district or my principal or parents to think that just because they come into fifth grade, I can remediate from fourth to what our district wants us to do in fifth and get them ready to pass sixth grade proficiency. How can you do it?"

Teacher V also offered valuable insight into what teachers may believe by suggesting that Ohio's testing should align to what she does in the classroom.

"I think that they have some things that they are questioning pretty accurately, but in other areas I think that they have not aligned with what our curriculum is in the classroom. I think that is really confusing for the children. Especially in the content area that we cover in the fifth grade, we spend a lot of time trying to make sure that they have a good background going into… the future in the sixth grade. Some of the questions that we can tell will be on the other proficiency tests have absolutely nothing to do with what we have been working so hard on. That just does not seem fair to me. So my biggest gripe is that the test is not aligned to our teaching. The curriculum that we teach does not line-up with what the children are tested on the proficiency tests."

Teacher Z, who comes closest to aligning with the content, processes, and format/style of the OPT, offered the most positive reaction to the researcher's questions about the OPT and whether or not it is testing the right things.

"I feel that it is here to stay no matter what. We might as well be doing our best to get the kids prepared for it…The science concepts that are on the sixth grade proficiency tests are good concepts, so I think the questions are good. In the paragraphs that they have to read, it has a lot of information, they just have to pick out the correct answer from that and read figures and discern what the answer is from looking at those."

Throughout the analysis of the transcripts, other lesser themes emerged that provide some insight their assessment strategies. For example, when asked

to comment on the use of extended response items and rubrics, only three teachers mentioned that they use rubrics. A fourth teacher (V) offered a curious response; "*In the back of the supplemental books there are some rubrics. I have noticed that in the later editions of the supplemental books there are more rubrics*," but she did not state that she uses them to assess student responses or projects. This data is supported by the document data, which found only four teachers submitting rubrics as requested by the researcher. However, of the eight rubrics that were submitted, all of them were fairly vague and lacking any real descriptive criteria for assessing the students' work. For example, most of the rubrics relied heavily on "criteria" such as these three separate items -- *organized*, *all necessary topic areas on poster*, *neatness*, and *teamwork*. For a seed collection project, another rubric offered *one point for every seed collected up to 10 points*, *five points for student's name and section*, and *five points for neatness*. Only one of the criteria with this rubric addressed science content or processes and it was a very low-level of cognition - *one point for each seed that was labeled*.

Another minor theme that emerged from analysis of the transcript data of the third grade teachers was that the reading level for the OPT is perceived to be too high for their students to succeed. Teacher D noted that, "*The OPT is more of a reading test than anything else is. I feel that the readability level is too hard for our students. I would not say in all cases but in a lot of cases the readability of the whole test is difficult.*" Teacher E's objection was that students "*have to be a reader to do the science. So a student who is not a good reader but knows their science knowledge, never gets it expressed because they are trapped by*

the fact that they can't read the test." While teacher C thought that the OPT *"is testing the right things,"* she found it "*difficult because it is mostly reading. If the children have difficulty in reading, you know they get very frustrated. My readers will do better on the OPT and my lower readers have a hard time."*

Summary of findings

This section summarizes primary, secondary, and ancillary findings.

Primary findings. This study produced three findings from analyses of Ohio's fourth and sixth grade science proficiency tests for years 2000 and 2001. First, multiple-choice items made up about three fourths of the total items from both years of both versions of the science OPT. These items were worth 55% of the possible total points. Second, the multiple choice questions were categorized into three groups - simple, reading, and figures/graphs/tables. The multiple-choice items requiring students to *process information from a figure, graph, or table* prior to choosing the correct response dominated all other forms of items including short and extended responses. Third, also based on analyses of both tests, the ability to *process scientific knowledge* is an essential skill for succeeding on the tests.

The researcher gleaned four primary findings from his comparison of the participants' data with his analysis of the OPT. 1) Comparison of the OPT data with the formal summative classroom assessment practices of the participants found little or no similarities. With the exception of one teacher participant, little evidence of alignment with any of the three categories, format and style, scientific process, and content, was found in the participants' document data. 2) Results of the document data for *scientific process* revealed that the tests used in the

participants' classroom assessments contained mostly low-level in terms of cognitive function. The reliance on low-level items occurred more frequently with the third grade teachers. 3) Analysis of the document data found that seventy-one percent (71%) of all of the participants' items were not aligned to the style and format found on the OPT. Teacher items generally fell into three categories - true/false, completion, and matching. The remaining items were predominantly low-level multiple-choice items requiring simple recall of information. 4) The teachers relied mostly on black-line masters with almost no use of teacher-designed assessments. During the interview sessions, eighty percent (80%) of the participants admitted relying mostly on black-line and publishers' tests.

Secondary findings. Some incongruities between the teacher's self-perceptions and their documented practices were also found. According to the transcript data, all of the participants believed that their instructional objectives were aligned to their formal assessments. However, only half of them believed that their curriculum was aligned to the outcomes assessed by the OPT. Second, according to the transcript data, all of the teachers felt that using the same format and style as the OPT will help students succeed on the OPT, yet the document data showed almost no alignment between analyses of their items and analysis of the OPT items. Third, the teachers' in this study believed that the content being assessed by their questions was aligned to the learning outcomes being assessed by the OPT. Eight of the ten teachers reported assessing some of the learning outcomes at least informally. However, the document data showed only a handful of items related to the content assessed by the OPT. Finally, since the teachers were focused on the life science stand during the

study, the researcher performed an in-depth study of this strand. For an entire quarter of formal assessments, the participants had only a very small number of items that assessed the same life science learning outcomes assessed by the OPT. Four examples of teachers with frequencies higher than three items that assessed the same life science learning outcomes as the OPT were found in the data.

There was some congruence between analysis of the document data and the participants' perceptions as found in analyses of the interview data and questionnaire data. For example, the fifth grade teachers perceived that they were assessing their science students at a higher level of cognition than did the third grade teachers. This finding was congruent with the document data.

Four additional findings from this investigation into the science assessment practices of third and fifth grade teachers are noteworthy. First, four teachers submitted rubrics for assessing short and extended response items or projects. The rubrics were generally very vague and lacked any real descriptive criteria for assessing their students' work. Second, sixty percent (60%) of the participants perceived that the textbook's tests on which they relied so demonstratively were aligned to the OPT, contrary to the findings of the researcher. Third, during the interviews, sixty percent (60%) of the participants expressed negative feelings for the OPT. Finally, eighty (80%) of the teachers in this study demonstrated a fairly weak level of knowledge about the OPT.

Ancillary finding. An ancillary discovery during this investigation occurred during the researcher's analysis of the 4th and 6th grade proficiency tests developed by a private publishing company. While attempting to code the

second level grains for the category of scientific process, the researcher encountered some ambiguity. Phone contacts with consultants from the Ohio Department of Education revealed mixed messages regarding the department's intent to link the items of the tests to the grains identified in their published *Information Guide*. In conversations with three different ODE consultants, each verified that these smaller grains were either "never linked with the test items" (Consultant A, 2002) or that there was "never any documentation linking these smaller grains with the test items" (Consultant B, 2001; Consultant C, 2001). However, Consultant B (2001) admitted that in the bidding process, the publishing company was given these same *Information Guides* used by the researcher as specifications for satisfying the contract between the Ohio Department of Education and the publishing company. Finally, the consultants were unwilling to develop this link when requested by this researcher. Consultant A (2001) admitted, "There really is no direct correlation between these smaller grains [as described in the assessment guides] and the tests…you [the researcher] should share this [revelation] with someone else in the department who has been here longer than I have."

Chapter 5 - Conclusions and Discussion

The final chapter of this investigation into the assessment practices of ten elementary science teachers discusses the significance and implications of the findings of this research. The discussion includes the relationship of the findings to previous research. The discussion is followed with recommendations for all levels of public educators. The chapter concludes with a delineation of the limitations of this study.

Significance of Findings, Implications, and Relationship with Previous Research

The lack of alignment between the teachers' assessments and the OPT is significant because it is widely accepted that teachers teach what they test. This premise suggests that the lack of alignment found in this study between the

teachers' assessments and Ohio's proficiency tests suggests a lack of alignment between their instruction and the objectives assessed by the OPT. In addition, the general lack of alignment to the style and format found on the OPT is a concern because the teachers and other researchers (Indiana Center for Evaluation, 2001; Haynie, 1994; Haynie, 1990a, 1990b, 1991, in press; Nungester & Duchastel, 1982; Smith, 2000; Borger et al, 1996; Hanny, 2001) believe that using the same format and style with their assessments will improve student performance on tests such as the OPT. Unfortunately, teacher items generally fell into three categories - true/false, completion, and matching. This result is a concern because these kinds of items do not appear on the OPT. The remaining items on their tests were predominantly low-level multiple-choice requiring simple recall of information. The result is a concern because the greatest portion of the OPT consists of higher-level multiple-choice items that require students to interpret and/or analyze figures, tables, and graphs.

Not only is it a concern that the appropriate content is most likely not being experienced by the students, the finding that teachers in this study were assessing students at a lower level of cognitive function than the items analyzed from the Ohio Proficiency Tests is a concern on two fronts. First, it would suggest that testing at this lower level fails to prepare students for the level of questioning that they will encounter on the state-mandated test. Other research suggests that lofty instructional goals designed to encourage understanding and critical thinking are undermined by tests that emphasize recognition of factual details that are inconsistent with the goals (Bol & Strage, 1996; Linn, 1990). Second, based on the premise that teachers teach to the level and content of their tests (O'Day &

Smith, 1993; American Association for the Advancement of Science, 2001), it would appear that the level of instruction and content used by the teachers in this study is not properly aligned with the level of learning and content being assessed by Ohio's tests to effectively facilitate positive results on the OPT. A similar concern is found in Bishop's study (2001) which suggests that the use of exams that test lower order items such as facts and procedures, are likely to result in too much class time being devoted to practicing low-level skills.

The document analysis revealed that the teachers relied heavily on the pre-packaged assessments (black-line masters) that accompany most textbook adoptions. In addition, there was a noticeable absence of teacher-designed assessments. This discovery was supported by the interview data with eight of the participants admitting to relying mostly on black-line and publishers' tests for their formal assessment of student learning. This discovery differs from the findings of previous researchers (Haynie, 1983, 1991, 1992; Herman & Dorr-Bremme, 1982; Mehrens, 1987; Mehrens & Lehmann, 1987; Newman & Stallings, 1982) who found most teacher assessments were done with teacher-made tests.

Some incongruities found during triangulation of the data suggest a level of dissonance at work between the perceptions of the teachers and their practices. First, the transcript data revealed that all of the participants believed that their instructional objectives were aligned to their formal assessments; yet, in contrast, the questionnaire data suggests that only half of them believed that their curriculum was aligned to the outcomes assessed by the OPT. Second, all of the teachers felt that using the same format and style as the OPT will help

students succeed on the OPT while the document data showed almost no alignment between analyses of their items and analysis of the OPT items. Third, the teachers in this study believed that the content being assessed by their questions was aligned to the learning outcomes being assessed by the OPT. Eight of the ten teachers reported assessing some of the learning outcomes at least informally. However, the document data showed only a handful of items related to the content assessed by the OPT. These two findings suggest that the teachers may not have a complete understanding of what content the OPT assesses and how it is assessed.

Secondary findings of this research convey additional implications. As noted, the rubrics used by less then half of the participants were vague and lacked any real descriptive criteria for assessing their students' work. The lack of use of proper rubrics for assessing short and extended response items or projects points to the need for inservicing teachers on the development and use of rubrics aligned to the OPT.

The relatively high level of expressed negative feelings for the OPT is a concern because it suggests that some teachers consciously or unconsciously may be resisting strategies for helping students succeed on the OPT. It may also explain why so many of the teachers in this study demonstrated a fairly weak level of knowledge about the OPT. These findings suggest that the teachers do not have enough of an understanding of the OPT and their own assessment strategies to develop a valid alignment between their assessment practices and the assessment format of the OPT. Just as important, this finding would suggest that many of the participants are incapable of reflecting on their personal

assessment practices because they lack sufficient information.

The revelation that the Ohio Department of Education never intended to link the items of the tests to the *grains* identified in their published *Information Guide* seriously concerns this researcher. It indicates a level of dysfunction that hurts the credibility of the state's system of accountability and contributes to the lack of teacher buy-in to the process which, in turn, leads to some of the resistance displayed by teachers as evidenced in newspaper editorials and witnessed by this researcher during his program of studies. This finding also lends credence to the findings of the researchers with Project 2061 (American Association for the Advancement of Science, 2001) and Webb (1999) who found inconsistency of linkage between the standards and the actual tests in their analysis of various states' high-stakes assessments. Webb (1999) found that, "A high percentage of standards and assessments (in science and math)...failed to achieve depth-of-knowledge consistency. In general, too high a frequency of items was below the depth-of-knowledge level of the corresponding objectives for there to be alignment." In Webb's study, reviewers analyzed the alignment of assessments and standards in mathematics and science from four states at a four-day institute. Six reviewers compared the match between assessment items and standards in mathematics and seven compared the match in science. Data from these analyses were processed and used to judge the degree of alignment on four criteria: categorical concurrence, depth-of-knowledge consistency, range-of-knowledge correspondence, and balance of representation. The analyses indicated that the standards of the four states varied in what content students were expected to know, the level of specificity at which expectations were

expressed, and organization. Nearly all of the sixteen assessment instruments reviewed incorporated some constructed-response items. Only one mathematics assessment for grade 10 from one state consisted solely of multiple-choice items. The items in three science and two mathematics assessments analyzed from one state were evenly divided between multiple-choice and constructed-response items. Assessments from the other three states included from 80% to 90% multiple-choice items. Alignment between assessments and standards varied across grade levels, content areas, and states without any discernable pattern. Assessments and standards of three of the four states satisfied the categorical concurrence criterion. This criterion, the most common conception of alignment, required the assessment and standards to include the same content topics. Alignment was found to be the weakest on the depth-of-knowledge consistency and range-of-knowledge correspondence criteria. Generally, assessment items required a lower level of knowledge and did not span the full spectrum of knowledge as expressed in the standards. However, for the knowledge and skills identified in the standards and addressed by the assessments, generally the assessment items were evenly distributed.

The lack of a clear understanding by teachers of how the OPT assesses their students has additional implications for policy makers and Ohio's Department of Education. The work of the Project 2061 researchers has suggested that many states' standards, especially in the content area of science, are not clear and are not aligned with their mandatory high-stakes assessments. This study and the work of Project 2061 suggest that the clarity of standards and the alignment of these kinds of tests may play an important role in the teachers'

lack of understanding. Providing clearly articulated standards for all educators and parents must be a high priority.

Discussion

Why would the participants' assessment and, potentially, their instruction be so poorly aligned with what is demanded of the students by Ohio's mandated testing program? It could have been the effect of a misaligned curriculum or teacher deviation from the planned curriculum. The participants' perceptions that the textbook's tests on which they so demonstratively relied were aligned to the OPT, provides another concern on a very deep level. This finding indicates that the textbook assessments are either not aligned with the textbook content, which is not likely, or the textbook is not aligned with the curriculum, which is more likely. It suggests that the black-line masters or publisher's assessment packages are aligned to textbooks which are not aligned to the content being assessed by OPT.

One possible explanation for the misalignment of the publishers' assessment packages and textbooks comes from researchers (American Association for the Advancement of Science, 2001) who have determined that most textbooks are not aligned with most states' large-scale assessments especially in the area of science. Project 2061 evaluated textbooks and their alignment with states' standards. Their goal is to produce resources which will be useful to commercial developers and publishers of instructional and assessment materials as well as districts and states that select and administer large-scale testing programs in addition to classroom teachers who create or assemble their own quizzes or tests. Andrew Ahlgren, associate director of Project 2061,

articulated the team's vision. "On the one hand, we intend that our assessment analysis procedure will contribute to writing better tasks and tests. On the other hand, we hope to change educators' views on what to expect of assessment tasks and help them make better choices among what publishers offer them (American Association for the Advancement of Science, 2001, Putting tests to the test, p. 1)."

That teachers may be relying so heavily on textbooks for designing lessons and assessments is of major concern. At a February 2001 conference sponsored by the American Association for the Advancement of Science (2001), education stakeholders agreed that the formula for teaching science to U.S children must be rewritten to improve student performance. "Science textbooks are an influential part of that formula. Poorly focused textbooks are a significant contributor to poor student performance, especially in the absence of good alternative materials therefore, teachers who must depend solely upon textbooks are unable to provide their student with a quality science education (p. 3)." During the conference, Ellen Standafer, vice president of science product development for one of the major science textbook publishers, offered reasons why the publisher's cannot align their books to the various states' standards. "The development of material requires that publishers serve an ever-growing number of special interest with different agendas. Add the economic limitations of custom publishing to the problem, and the challenge for publishers becomes finding creative ways to integrate diverse and, sometimes, conflicting standards. Publishers are further challenged to bridge the gap between what teachers want for their classrooms and what their state or district mandates. We look for as

much commonality in standards as possible. Obviously, these decisions are impacted by market size (p. 3)." It, therefore, comes as no surprise to this researcher that Texas and California drive the publishing process.

Regardless of what the participants in this study believed about the design of their tests and the quality of the publisher's tests, in practice, these kinds of assessments appear to be poorly constructed and demand low levels of cognitive processing. Why were the teachers' tests in this study so poorly aligned? Perhaps other studies provide some insight. Research on teachers' assessment practices have shown that many teachers are inadequately trained and ill-prepared to develop, administer, and interpret the results of various types of assessments (Daniel & King, 1998; Plake, Impara, & Fager, 1993; Wise, Lukin, & Roos, 1991), including alternative assessments (Stiggins, 1992; Wiggins, 1991). Bol and her colleagues (1998) found that the frequencies with which teachers used different methods of assessment were related to their perceptions of how well prepared they were to develop and administer the assessments. Is time a factor? Do teachers fail to develop more of their own assessment items because they choose to place their time into planning instruction? These are questions that might be worthy of additional research.

Recommendations

This section discusses recommendations for all levels of educators in addition to policymakers and textbook publishers.

Teachers. Based on the findings of this investigation, the researcher proposes that teachers consider different kinds of assessments as parts of a continuum when attempting to align their assessments with the purpose and

goals of their instruction. In order from left to right on the continuum, these items would consist of: *true-false*, *fill-in-the-blank*, *modified true-false*, *matching*, *multiple-choice* (simple, text, and figure/graph/table), *simple student-constructed responses*, *extended responses*, *directly observed responses*, *simple demonstration*, *comprehensive demonstration*, and *collections of work over time*. These items can be spaced on the continuum relative to their application as meaningful tools for assessing certain kinds of student learning. The continuum shown below is based partially on Elliot's (1995) conceptualization of *performance assessment formats*, which he suggests ranges from the simplest student-constructed responses to comprehensive demonstrations or collections of work over time. The researcher has modified Elliot's model by adding the items shown to the left of Ellliot's *performance assessments* and delineating the three kinds of multiple-choice items found on the OPT.

The left side of the continuum contains those assessments that require students to select a response. The right side of the continuum consists of assessments in which students construct their response. The science portion of the OPT is represented as a continuum ranging from strictly content-oriented test items to process-oriented and extension items. When the *continuum of the science OPT* is overlaid with the *assessment continuum* and Elliot's *performance assessment continuum*, it is the writer's belief that the OPT and the performance assessment continuums overlap with the short and extended response items of the both the OPT and Elliot's continuum. However, both leave a gap with the extreme left portion of the researcher's newly conceptualized *total assessment continuum*. In other words, neither of the continuums covers the true/false, fill-in, modified true/false, or the matching assessments. Whether or not the classroom teacher should use the uncovered portions of this continuum when assessing student learning is best determined by considering the level of learning that the item or assessment instrument is attempting to measure. To identify the level of learning, teachers would be well served by using a proven taxonomy such as Bloom's (1956). Skills and processes above the level of knowledge would be best assessed using assessments from the OPT and performance assessment portions of the continuum.

Finally, teachers should consider that well-designed tests serve to motivate and help students structure their learning. Crooks (1988), McKeachie (1986), and Wergin (1988) suggest that students study in ways that reflect how

they think they will be tested. If they expect an exam focused on facts, they will memorize details. If they expect a text that will require problem solving and processing of information, they will work toward these higher levels of cognitive function.

Administrators. To meet the needs of teachers identified by this study, it is recommended that school districts use some of their staff development funds as stipends for teachers to write assessment items aligned to the OPT during summer workshops. Teachers in these workshops would work collaboratively with each other and experts in psychometrics to design a bank of assessment items for their curriculum objectives. This researcher believes that students will ultimately improve their achievement on statewide tests if their school's learning objectives are properly aligned to the state's standards and the classroom assessments are valid, reliable, and aligned with the format/style and level of learning of the statewide assessments. In addition, students will also improve their achievement because teachers will be more deliberate and focused with the purpose and methods used with their instruction if they have a predetermined set of assessments and items from which to choose. It is also important that in developing these assessments and items, teachers should keep in mind that going further on the assessment continuum for objectives that are on the higher end of learning (processing, extending, etc) will be a positive force in driving the instruction to a higher level of learning which, again, should result in improved student achievement.

Textbook adoption committees should enter the process cognizant of the shortfalls of pre-packaged assessments and the poor alignment most textbooks

have with many state's standards. Realizing that textbooks should only be one source of a multitude of resources is an important first step in the process of adopting any textbook. Teachers should be encouraged by administrators to look beyond the textbook when developing lessons aligned to state standards. They should consider "the world" as an unlimited source of materials for implementing the curriculum and not feel restricted to the adopted textbook and its accompanying "bells and whistles".

School leaders should commit to using assessment information to improve instruction by providing time for their teachers to meet, discuss, and make instructional decisions based on data. Judgments about the efficacy of the district's curriculum and instructional practices including the amount of time allocated for various content areas should be made on a continuous basis. However, because state assessments are typically infrequent, more frequent and effective classroom assessment must be facilitated and encouraged by administrators because frequent assessment has been shown to be more effective in gauging student progress and it is associated with increased achievement.

Finally, school leaders must insure that the teaching of science includes the opportunity to apply and extend scientific knowledge performing authentic tasks in a laboratory setting at all grade levels. The laboratory setting need not be patterned after the traditional high school lab nor should it be confined to the classroom. But as Nelson (2001) suggests, "If textbooks, tests, and classroom activities are to focus on helping students gain an in-depth understanding of the ideas and skills that are most important for science literacy, they need to

incorporate what is known about how people learn science and how to probe for conceptual understanding and higher order skills (p 2)." According to Dietel et al, (1991), evidence about the nature of learning suggests that instruction which strongly emphasizes structured drill and practice on discrete, factual knowledge does students a disservice. Learning isolated facts and skills is more difficult without meaningful ways to organize the information and make it easy to remember. Applying those facts and skills to the kinds of processing and extension tasks demanded by the majority of the items on the fourth and sixth grade OPT becomes a separate and difficult task without the real world learning that effective laboratory experiences can provide students. Educational leaders must provide teachers the proper resources for this kind of learning. Finally, policymakers must require that elementary teacher preservice preparation for licensure include adequate preparation for the teaching of science through the integration of effective laboratory experiences. If science is included in states' high-stakes testing requirements for elementary students, then licensure of elementary teachers should include this kind of requirement.

State Departments of Education. As mentioned previously, Project 2061 researchers have suggested that many state's standards are not clear and are not aligned with their mandatory high-stakes assessments. Providing clearly articulated standards for all educators and parents must be a high priority of each state that uses these kinds of tests to rate schools and students. To do otherwise, is unjust to students, educators, and the communities they serve and borders on impertinence.

Once the standards are clear and the tests are aligned to them, the work

has only begun. Appropriate and adequate professional development must be provided through intermediate education agencies and the local school districts. Current professional development on assessment to provide teachers with information on how to build better tests is lacking. Knowledgeable consultants must be made available to these agencies on a continuous basis. Emphasis on training that includes basic item development, how to score short and extended items with a rubric, how to design effective rubrics, and linking test items to the curriculum would improve teachers' abilities to better understand what resources and strategies were the most effective and whether students mastered the learning. As new teachers enter the field, they will need the same training. Training opportunities will need to be made available on a year-round basis to accommodate the wide-ranging differences in school calendars. Expenses for training and stipends for teachers to attend the training during time outside of teachers' normal contracted hours should not be the burden of the individual schools or it becomes another unfunded mandate, or worse, inequitable for those districts too poor to fund it.

Policymakers. Legislation mandating that state education agencies establish clear learning standards should be passed. Learning standards should be clear in terms of the expectations for what students are required to know and do with minimal ambiguity. Legislators should require all state-mandated assessments be aligned to the standards. Finally, policymakers at all levels must provide adequate funding for experts to provide high quality training experiences in the areas of classroom assessment, interpreting assessment data, school-improvement planning, and curriculum mapping and alignment.

Limitations

This section describes the limitations of the methods used in this study and concludes with considerations for future research.

Local Bias. The study was conducted with participants from one school district and, therefore, should not be generalized to include all teachers. In addition, all schools in Ohio do not universally use the textbook series used by the participants in this study.

Limited Sample. Although the participants in this study cut across several experience levels and all teachers from the targeted grade levels of the district participated, the population was gender specific and somewhat skewed to the higher experience levels. In addition, the sample contained only ten teachers (five from third grade and five from fifth grade).

Reliability of the Questionnaire Items. The questionnaire used by the researcher, although useful for triangulating with the other sources of data, was not field tested for reliability. Researchers and practitioners interested in using this tool should consider making refinements after additional field-testing to determine its reliability and validity with regard to their unique situation. Despite these limitations, the basis for the tool has practical implications for teachers and administrators interested in assessing alignment factors with their curriculum, instruction, and assessment with regard to state standards.

Reliability of Researcher's Analysis of Process. The use of the *grains* to classify the subsets of the three major scientific processes was based on the researcher's interpretation of the meaning of each of the terms. When contacted by the researcher, ODE consultants were unwilling to assist with this process.

While the same researcher conducted the analysis of both the OPT and the teachers' assessment items to maintain the relative integrity of the comparisons between the OPT and the teachers' data, it is recognized that a different analyst could arrive at different results. If future investigators wish to use this method to analyze the subsets of the three scientific processes, it is recommended that a team of two or more analysts be used to classify the items according to these grains in order to establish a statistical level of confidence through the calculation of a reliability coefficient.

Reliability of Researcher's Analysis of Content. To determine which *learning outcome* the participants' items were assessing, the researcher used the descriptions of the learning outcomes provided by ODE. While this information was available in ODE's Answer Key, researcher's judgment was used to determine which learning outcome the participants' items were assessing. By conducting the analysis of the OPT first, the researcher was able to see precise models for each of the *learning outcomes*. If future investigators wish to use this method to analyze the content of teachers' assessment items, it is recommended that a team of two or more analysts be used to classify the items according to learning outcomes in order to establish a statistical level of confidence through the calculation of a reliability coefficient.

Future Considerations

This study answered the original research questions posed by the investigator. However, its findings have generated additional questions. What is the level of alignment between the assessment practices of other teachers in other schools and their respective state tests? If there are alignment issues, what

is/are the principal cause(s)? Is it poor preservice and inservice training with effective test construction or is it that teachers do not have enough time to develop reliable and valid tests that are aligned to their state's tests? Or, is the problem a result of misalignment of the curriculum? Perhaps, the problem is more related to a general lack of understanding of or a bias against the states' tests by many teachers than any other cause. The current trend with high-stakes testing and holding schools accountable for student success on these tests does not appear to be to have an end. The legislators and the general public have given a clear mandate to public school educators at all levels. It would appear that additional research in this area is warranted. Finally, it is recommended that teachers, with the help of qualified administrators or outside experts, conduct action research to sample their assessments with the purpose of providing meaning-making data. Despite the negative implications of this study, evidence that progress is being made is found in the transcript data of one of the participants: *"I think as a teacher that the only way our students will be able to pass the proficiency is if we have the goals and objectives linked with the proficiency and that is what we work towards."* The researcher remains hopeful that, if the states' clearly define the targets (standards) and teachers and administrators focus their attention on them, the targets will be hit.

References

American Association for the Advancement of Science. (2001). Project 2061. Collaborating to create better tests: A publisher's perspective. 2061

Today, Spring/Summer, 2, 1, 3-5.

American Association for the Advancement of Science. (2001). Project 2061. Putting tests to the test. 2061 Today, Spring/Summer, 2, 1, 1-2.

Applebee, A. (1981). Writing in the secondary school: Current practice in English and the content areas (Research Report No. 21). Urbana, IL: National Council of Teachers of English.

Applebee, A. (1993). Literature in the secondary school. Urbana, IL: National Council of Teachers of English.

Asp, E. (1998). The relationship between large-scale and classroom assessment. In R. Brandt. (Ed.), Assessing student learning: New rules, new realities (pp. 17-46). Arlington, VA: Educational Research Service.

Bishop, J. H. (2001). A steeper, better road to graduation. Education Next, Winter, online at www.educationnext.org/20014/56.html, December 3, 2001.

Black, P. and William, D. (1998). Assessment and classroom learning. Assessment in Education. March, 7-74.

Bloom, B. (1956). Taxonomy of Educational Objectives. Vol. 1: Cognitive Domain. NY: McKay.

Bloom, B. (1984). The search for methods of group instruction as effective as one-to-one tutoring. Educational Leadership, 41, 4-17.

Bol, L., Stephensen, P. L., Nunnery, J. A., & O'Connell, A. A. (1998). Influence of experience, grade level, and subject area on teachers' assessment practices. Journal of Educational Research, 91(6), 323-330.

Bol, L., & Strage, A. (1996). The contradiction between teachers' instructional

goals and their assessment practices in high school biology courses. Science Education, 80, 145-163.

Bol, L., Nunnery, J., Stephenson, P., Mogge, K. (2000). Changes in teachers' assessment practices in the new American schools restructuring models. Teaching and Change, Thousand Oaks: CA.

Brookhart, S. M. (1994). Teachers' grading: Practice and theory. Applied Measurement in Education, 7(4), 279-301.

Brophy, J. E. & Alleman, J. (1991). Activities as instructional tools: A framework for analysis and evaluation. Educational Researcher, 20, 9-23.

Burden, P. R., & Byrd, D. M. (1994). Methods for effective teaching. Boston: Allyn & Bacon.

Campbell C, and Evans, J. A. (2000). Investigation of preservice teachers' classroom assessment. The Journal of Educational Research (Washington, D.C.) 93, 6, 350-5.

Borger J., Cano, M., Collins, C., Evans, W., Perlman, C., Qualls, J., Wood, J.. (1996). Preparing your elementary students to take standardized tests. Office of Accountability, Chicago Public Schools. Chicago School Reform Board of Trustees: Chicago.

Cizek, G. J. (2000). Pockets of resistance in the assessment revolution. Educational Measurement: Issues and Practice, summer, 16-33.

Consultant A, (2002) Phone conversation with a science consultant from the Ohio Department of Education, November 12. (A release form was not secured prior to the conversation; therefore the consultant's identity remains protected.)

Consultant B, (2001). Phone conversation with Ohio Department of Education consultant, School Standards and Assessment, November 12. (A release form was not secured prior to the conversation; therefore the consultant's identity remains protected.)

Consultant C, (2001). Phone conversation with Ohio Department of Education science consultant, November 12. (A release form was not secured prior to the conversation; therefore the consultant's identity remains protected.)

Cozby, P. C. (1997). Methods in behavioral research. Mountain View, CA: Mayfield.

Cromey, A. (2000). Using student assessment data: What can we learn from schools? Policy Issues, North Central Regional Educational Laboratory.

Crooks, T. J. (1988). The impact of classroom evaluation practices on students. Review of Educational Research, 58, 438-481.

Curriculum Alignment Project, (2000). Tier I and Tier II Curriculum Alignment Project for the *High School Graduation Qualifying Exam* sponsored by the Ohio Department of Education. October 11 memorandum from the office of the Regional Title One Facilitators, Region Four, Dayton, OH.

Daniel, L. G., & King, D. A. (1998). Knowledge and use of testing and measurement literacy of elementary and secondary teachers. Journal of Educational Research, 91(6), 331-344.

Dietel, R. J., Herman, J. L., and Knuth, R. A. (1991). What does research say about assessment? North Central Regional Educational Laboratory, Oak

Brook.

Doyle, W. (1983). Academic work. Review of Educational Research, 53,159-199.

Elliot, S. (1995). Creating meaningful performance assessments. ERIC Digest E531. Derived from "Creating meaningful performance assessments: Fundamental concepts", S. N. Elliott, see ED 375 566.

Fleming, M., & Chambers, B. (1983). Teacher-made tests- Windows on the classroom. In W. E. Hathaway (Ed.), New directions for testing and measurement: Vol. 19. Testing in the schools (pp. 29-38). San Francisco: Jossey Bass.

Frary, R. B., Cross, L. H., & Weber, L. J. (1993). Testing and grading practices and opinions of secondary teachers of academic subjects: Implications for instruction in measurement. Educational Measurement: Issues and Practice, 12(3), 23-30.

Gallagher, J. J. (1991). Prospective and practicing secondary school science teachers' knowledge and beliefs about the philosophy of science. Science Education, 75,121-133.

Goodlad, J. I. (1984). A place called school. New York: McGraw-Hill.

Gronlund, N. E., & Linn, R. L. (1990). Measurement and evaluation in teaching (6th ed.). New York: Macmillan.

Guidelines for the fourth grade proficiency tests, (n.d.) Ohio Department of Education. Online, http://www.ode.state.oh.us/proficiency/sample_tests/fourth/4iscie.pdf

Guidelines for the fourth grade proficiency tests, (n.d.) Ohio Department of Education. Online, http://www.ode.state.oh.us/proficiency/sample_tests/

sixth/6iscie.pdf).

Gullickson, A. R., & Ellwein, M. C. (1985). Post hoc analysis of teacher-made tests: The goodness-of-fit between prescription and practice. Educational Measurement: Issues and Practice, 4(1), 15-18.

Haertel, E. (1986). Choosing and using classroom tests: Teachers' perspectives on assessment. Paper presented at the annual Meeting of the American Educational Research Association, San Francisco.

Haertel, E. H. (1999). Performance assessment and educational reform. Phi Delta Kappan, 662-666.

Hanny, R. J. (2001). Teacher-made tests and the VA SOL. On-line, wysiwyg://16http:www.pen.k12.va...OE/Instruction/wmstds/solass.shtml, June 19.

Haynie, R. J. (1983). Student evaluation: The teacher's most difficult job. Monograph Series of the Virginia Industrial Arts Teacher Education Council, Monograph Number 11.

Haynie, R. J. (1990a). Effects of tests and anticipation of tests on learning via videotaped materials. Journal of Industrial Teacher Education, 27(4), 18-30.

Haynie, R. J. (1990b) Anticipation of tests and open space laboratories as learning variables in technology education. In J. M. Smink (Ed.), Proceedings of the 1990 North Carolina Council on Technology Teacher Education Winter Conference. Camp Caraway, NC:NCCTTE.

Haynie, R. J. (1991). Effects of take-home and in-class tests on delayed retention learning acquired via individualized, self-pace instructional texts.

Journal of Industrial Teacher Education, 28(4), 52-63.

Haynie, R. J. (1992) Post hoc analysis of test items written by technology education teachers. Journal of Technology Education, 4(1), 27-40.

Haynie, W. J. (1994). Effects of Multiple-choice and short-answer tests on delayed retention learning. Journal of Technology Education, 6(1).

Herman, W. J. & Dorr-Bremme, D. W. (1982). Assessing students: Teachers' routine practices and reasoning. Paper presented at the annual meeting of the American Educational Research Association, New York.

Herman, J. L., Aschbacher, P. R., & Winters, L. (1992). A practical guide to alternative assessment. Alexandria, VA: Association for supervision and Curriculum.

Indiana Center for Evaluation. (2001). Highlights from a case study of key effective practices in Ohio's improved school districts: report of results from phase I. Document presented to 2001 Ohio Statewide Administrators Conference: Strategies used by Ohio schools to increase achievement, August 6-7.

Jenkins, S. (2000). Conversation with Sandy Jenkins, Region *4 Title One Facilitator* assigned to academic watch and emergency level school districts, September 22.

Kahn E. A. (2000). A case study of assessment in a grade 10 English course. The Journal of Educational Research. (Washington, D.C.) 93 (5), 276-86.

Kane, M. B., & Khattri, N. (1995). Assessment reform: A work in progress. Phi Delta Kappan, 77, 30-32.

Linn, R. L. (1990). Essentials of student assessment: From accountability to

instructional aid. In S. Tozer, T. H. Anderson, & B. B. Armbruster (Eds.), Foundational studies in teacher education: An examination (pp. 124-138). New York: Teachers College Press.

Linn, R. L. (1993). Educational Assessment: Expanded expectations and challenges (CSE Technical Report). Los Angeles, CA: University of California. , Center for Research on Evaluation, Standards, and Student Testing.

Loveless, T. (2000). The Brown Center report on American education: 2000. How well are American students learning? Available online: http://www.brook.edu/GS/brown/bc_report/BC_Report_hp.htm

Madaus, G. R, Maxwell West, M., Harmon, M. C., Lomax, R. G., & Viator, K. A. (1992). The influence of testing on teaching math and science in grades 4-12: Executive summary. Center for the Study of Testing, Evaluation, and Educational Policy, Boston College: National Science Foundation.

Marso, R. N., & Pigge, F. L. (1991). An analysis of teacher-made tests: Item types, cognitive demands, and item construction errors. Contemporary Educational Psychology, 16, 279-286.

McDonnell, L. M. (1994). Policymakers' views of student assessment. (CSE Technical Report No. 378). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.

McKeachie, W. J. (1986). Teaching tips. (8th ed.) Lexington, Mass: Heath.

McMillan, J. H. (1997). Classroom assessment: principles and practice for effective instruction. Needham Heights, MA: Allyn & Bacon.

McNeil, L. M. (1986). Contradictions of control: School structure and school

knowledge. New York: Routledge & Kegan Paul.

Milton, O. Pollio, H. R. and Eison, J. A. (1986). Making sense of college grades: Why the grading system does not work and what can be done about it. San Francisco: Jossey-Bass.

Mehrens, W. A. (1987). Educational tests: Blessing or curse? Unpublished manuscript. 1987.

Mehrens, W. A. & Lehmann, I. J. (1987). Using teacher-made measurement devises. NASSP Bulletin, 71(496), 36-44.

Mehrens, W. A. (1989). Preparing students to take standardized achievement tests. Practical Assessment, Research & Evaluation, 1 (11). On-line, http://ericae.net/pare/getvn.asp?v=1&n=11

Nelson, G. D. (2001). Remarks on the release of the NAEP 2000 science assessment results. American Association for the Advancement of Science. On-line, November 20, 2001, http://www.project2061.org/newinfo/press/r10/1120a.htm

Newman, D. C. & Stallings, W. M. (1982). Teacher competency in classroom testing, measurement preparation, and classroom testing practices. Paper presented at the Annual Meeting of the National Council on measurement in Education, March (In Mehrens & Lehmann, 1987)

Nungester, R. j., & Duchasstel, P. E. (1982). Testing versus review: Effects on retention. Journal of Educational Psychology, 74(1), 18-22.

O'Day, J. A., & Smith, M. S. (1993). Systemic reform and educational opportunity. In S. H. Fuhrman. (Ed.), Designing coherent educational policy (pp. 250-312). San Francisco: Jossey-Bass.

Ohio Department of Education. (n.d.) Information Guide for the Fourth Grade Proficiency Test. Retrieved November, 2001 from http://www.ode.state.oh.us/proficiency/sample_tests/fourth/4iscie.pdf

Ohio Department of Education. (n.d.) Information Guide for the Sixth Grade Proficiency Test. Retrieved November, 2001 from http://www.ode.state.oh.us/proficiency/sample_tests/sixth/6iscie.pdf

Ohio Department of Education. (2000). Answer Key. Fourth Grade Proficiency Test. Available on-line, http://www.ode.state.oh.us/proficiency/00Tests/grade4/4sci00ans.pdf or by writing the Ohio Department of Education, Assessment Center, MS 507, 25 S. Front Street, Columbus, OH 43215.

Ohio Department of Education. (2000). Answer Key. Sixth Grade Proficiency Test. Available on-line, http://www.ode.state.oh.us/proficiency/00Tests/grade6/6sci00ans.pdf or by writing the Ohio Department of Education, Assessment Center, MS 507, 25 S. Front Street, Columbus, OH 43215.

Ohio Department of Education. (2001). Answer Key. Fourth Grade Proficiency Test. Available on-line, http://www.ode.state.oh.us/proficiency/01Tests/grade4/4sci01ans.pdf or by writing the Ohio Department of Education, Assessment Center, MS 507, 25 S. Front Street, Columbus, OH 43215.

Ohio Department of Education. (2001). Answer Key. Sixth Grade Proficiency Test. Available on-line, http://www.ode.state.oh.us/proficiency/01Tests/grade6/6sci01ans.pdf or by writing the Ohio Department of Education, Assessment Center, MS 507, 25 S. Front Street, Columbus, OH 43215.

Patton, M. Q. (1990). Qualitative evaluation and research methods. Newbury

Park, California: Sage.

Plake, B. S., Impara, J. C., & Fager, J. J. (1993, Winter). Assessment competencies of teachers: A national survey. Educational Measurement: Issues and Practice, 10-12.

Porter, A. C. (1995). Developing opportunity-to-learn indicators of the content of instruction: Progress report. Madison, WI: Wisconsin Center for Education Research.

Roeber, E. (1995). Emerging student assessment systems for school reform. ERIC Digest, ED389959, Greensboro, NC: Clearinghouse on counseling and Student Services.

Scharfer, D., & Lissitz, R. W. (1987). Measurement training for school personnel: Recommendations and reality. Journal of Teacher Education, 57-63.

Schmidt, W. H., & McKnight, C. (1995). Surveying educational opportunity in mathematics and science: An international perspective. Educational Evaluation and Policy Analysis, 3, 337-353.

Semb, G., & Spencer, R. (1976). Beyond the level of recall: An analysis of complex educational tasks in college and university instruction. In L. E. Farley & E. A. Vargus (Eds.), Behavioral research and technology in college and university teaching (pp. 115-125). Proceedings of the 3rd National Conference on Behavioral Research and Technology in College and University Teaching, Gainesville: Florida University Press.

Simmons, W., & Resnick, L. (1993). Assessment as the catalyst of school reform. Educational Leadership, 50, 11-15.

Smith, S. H. (2000). School wide test preparation: one elementary school's

instructional approach that dramatically raised standardized test scores. Arlington, VA: Educational Research Service.

Stenmark, J. K. (1991). Mathematics assessment: myths, models, good questions, and practical suggestions. Reston, VA: The National Council of Teachers of Mathematics, Inc.

Squire, J. R., & Applebee, R. K. (1968). High school English instruction today. Urbana, IL: National Council of Teachers of English.

Stiggins, R. J., Frisbie, D. A., & Griswold, P.A. (1989). Inside high school grading practices: Building a research agenda. Educational Measurement: Issues and Practice, 8(2), 5-14.

Stiggins, R. J. (1992). In teachers' hands: Investigating the practices of classroom assessment. Albany: State University of New York Press.

Thomas, J. W., Bol, L., Warkentin, R. W, Wilson, M., Strage, A., & Rohwer, W. D., Jr. (1993). Interrelationships among students' study activities, self-concept of academic ability, and achievement in high school biology courses. Applied Cognitive Psychology, 7,499-532.

Tyler, R. (1950). Basic principles of curriculum and instruction. Chicago: University of Chicago Press.

Webb, N. L. (1999). Research Monograph No. 18: Alignment of science and mathematics standards and assessments in four states, National Institute for Science Education, University of Wisconsin-Madison, Council of Chief State School Officers Washington, DC, August.

Wenglinsky, H. (2002). How teaching matters: Bringing the classroom back into discussions of teacher quality. A Policy Information Center Report. Educational Testing Service. Available on-line: http://www.ets.org/research/pic/teamat.pdf, 01-08-02.

Wergin, J. F. (1988). Basic issues and principles in classroom assessment. In J.H. McMillan (ed.), Assessing Students' Learning. New Directions for Teaching and Learning., no. 34. San Francisco: Jossey-Bass.

Wilson, L. W. (2002). Better instruction through assessment: what your students are trying to tell you. Larchmont, NY: Eye on Education.

Wiggins, G. (1991). Standard, not standardization: Evoking quality student work. Educational Leadership, 51, 74-79.

Wise, S. L., Lukin, L. E., & Roos, L. L. (1991). Teacher beliefs about training in testing and measurement. Journal of Teacher Education, 42, 37-42.

Wolf, D. P., LeMahieu, P. G., & Eresh, J. (1992). Good measure: Assessment as a tool for educational reform. Educational Leadership, 49, 8-13.

301 405-8134

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: Assessment Practices of Third and Fifth Grade Science Teachers: A comparison to the style/format, process, and content of Ohio's Proficiency Tests

Author(s): David Janson

Corporate Source:

Publication Date: May 22, 2002

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

Level 1

☐

Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) and paper copy.

The sample sticker shown below will be affixed to all Level 2A documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2A

Level 2A

☐

Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only

The sample sticker shown below will be affixed to all Level 2B documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2B

Level 2B

☒

Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, → please

Signature: David Janson

Printed Name/Position/Title: Dr. David Janson, Director

Organization/Address: Piqua City Schools

Telephone: 937 773-7321 FAX: 937 778 4517

E-Mail Address: djanson@piqua.org Date: 7-8-03

(over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, *or*, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

University of Maryland
ERIC Clearinghouse on Assessment and Evaluation
1129 Shriver Lab, Bldg 075
College Park, MD 20742
Attn: Acquisitions

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

University of Maryland
ERIC Clearinghouse on Assessment and Evaluation
1129 Shriver Lab, Bldg 075
College Park, MD 20742
Attn: Acquisitions